

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(54) International Patent Classification <sup>6</sup> :</b> <b>A61K 48/00, C07H 21/04, C12N 5/06,</b> <b>15/00, 15/11, 15/85, 15/86</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 96/40271</b> <b>(43) International Publication Date:</b> 19 December 1996 (19.12.96)
<b>(21) International Application Number:</b> PCT/US96/09430 <b>(22) International Filing Date:</b> 6 June 1996 (06.06.96)  <b>(30) Priority Data:</b> 08/473,845 7 June 1995 (07.06.95) US  <b>(71) Applicant:</b> YALE UNIVERSITY [US/US]; 451 College Street, New Haven, CT 06520 (US).  <b>(72) Inventor:</b> GLAZER, Peter, M.; 49 Thirsty Hill Road, Guilford, CT 06437 (US).  <b>(74) Agent:</b> KARTA, Glenn, E.; OncorPharm, Inc., 200 Perry Parkway, Gaithersburg, MD 20877 (US).		<b>(81) Designated States:</b> AU, BR, CA, CN, CZ, FI, HU, IL, JP, KP, KR, MX, NO, NZ, SG, SK, UA, Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> TREATMENT OF HEMOGLOBINOPATHIES  <b>(57) Abstract</b>  The present invention relates to management and treatment of hemoglobinopathies, such as sickle cell anemia and $\beta$ -thalassemia. The invention also relates to developing research animals and cell lines for the study of hemoglobinopathies and their therapies. The invention utilizes third strand oligonucleotides to target double-stranded nucleic acid sequences in or near the globin genes or in or near sequences controlling expression of those genes to cause either a desired mutation or nucleic damage to stimulate homologous recombination with a supplied donor nucleic acid.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

## TREATMENT OF HEMOGLOBINOPATHIES

5

### BACKGROUND OF THE INVENTION

#### 1. Field of the Invention

The present invention relates to management and treatment of hemoglobinopathies, such as sickle cell anemia and  $\beta$ -thalassemia. The invention also relates to developing research animals and cell lines for the study of hemoglobinopathies and their therapies. The invention utilizes third strand oligonucleotides to target double-stranded nucleic acid sequences in or near the globin genes or in or near sequences controlling expression of those genes to cause either a desired mutation or nucleic damage to stimulate homologous recombination with a supplied donor nucleic acid.

20

#### 2. Description of Related Art

##### Third Strands and Triple-Stranded DNA

Oligonucleotides (third strands) can bind to double-stranded DNA to form triple-stranded helices (triplexes) in a sequence specific manner.

Oligonucleotide-mediated triplex formation has been shown to prevent transcription factor binding to promoter sites and to block mRNA synthesis in vitro and in vivo (Blume et al., *Nucleic Acids Res.* 20:1777 (1992); Cooney, et al., *Science* 241:456 (1988); Duval-Valentin, et al., *Proc. Natl. Acad. Sci. USA.* 89:504 (1992); Grigoriev, et al., *Proc. Natl. Acad. Sci. USA.* 90:3501 (1993); Grigoriev, et al., *J. Biol. Chem.* 267:3389 (1992); Ing, et al., *Nucleic Acids Res.* 21:2789 (1993); Maher, et al., *Science* 245:725 (1989); Orson, et al., *Nucleic Acids Res.* 19:3435 (1991); Postel, et al., *Proc. Natl. Acad. Sci. USA.* 88:8227

(1991); Young, et al., *Proc. Natl. Acad. Sci. USA* 88:10023 (1991)). Such inhibition of expression, however, is transient, depending on the sustained presence of the oligonucleotides. It also depends on the stability of the triple helix, which can be disrupted by transcription initiated at nearby sites (Skoog and Maher, *Nucleic Acids Res.* 21:4055 (1993)). To overcome these problems, methods to prolong oligonucleotide-duplex interactions using DNA intercalating or cross-linking agents have been explored in experiments to block transcription initiation or elongation (Grigoriev, et al., *Proc. Natl. Acad. Sci. USA* 90:3501 (1993); Grigoriev, et al., *J. Biol Chem.* 267:3389 (1992); Sun, et al., *Proc. Natl. Acad. Sci. USA* 86:9198 (1989); Takasugi, et al., *Proc. Natl. Acad. Sci. USA*, 88:5602 (1991)).

Instead of transiently blocking gene expression, in the present invention third strands are used to target or direct mutagenesis or homologous recombination to specific sites in or near selected globin genes in order to produce permanent changes in gene function and expression. Long-term blocking of the DNA target is, therefore, not necessary. The fact that DNA damage and mutagenesis can be directed in this sequence specific manner by third strands is evidenced by mutagenesis "footprints" at the site resulting from unrepaired damage (Havre, et al., *Proc. Natl. Acad. Sci. USA* 90:7879 (1993); Havre and Glazer *J. Virology* 67:7324 (1993)).

#### The third-strand binding code and binding motifs

The third strand binding code dictates the sequence specificity for binding third strands in the major groove of double-stranded DNA to form a triple-stranded helix or triplex. Third-strand binding differs from the familiar Watson-Crick complementarity principle (A:T/U and G:C) for the double-stranded helix in two major respects: (1) the third-strand binding code is degenerate, and (2) third strands bind only to double-strands which contain a

sequence of adjacent (or run of) purine bases (A or G) in one of the strands, which here will be called the center or core strand. The third-strand binding code is illustrated in the chart below.

5

10

		Third	Strand	Base		
Center Strand Purine		A	U/T	I	G	C
A		+	+	+	-	-
G		-	-	+	+	+

In the center of the chart, a "+" means the bases are complementary or correspondent, and a "-" means they are not complementary or not correspondent. The bases are: A = adenine (purine); G = guanine (purine); C = cytosine (pyrimidine); T = thymine (pyrimidine); U = uridine (pyrimidine in RNA); I = inosine (purine, universal third-strand binding base)

20

Subject to the third-strand binding code, there are a number of "motifs" which further describe third-strand binding to purine center-strand targets. The motifs describe, for example, whether the third-strand must bind parallel or antiparallel to the center target strand (polarity); and to some extent the motifs describe center-strand sequence and nearest neighbor effects on binding.

25

#### Hemoglobins and Hemoglobinopathies

30

Only those aspects of basic hemoglobin biology, physiology, and terminology relevant to the present invention are discussed here. For a full discussion of hemoglobin and diseases related to abnormal hemoglobins (hemoglobinopathies) see Bunn and Forget ("Hemoglobin: Molecular, Genetic and Clinical Aspects", W.B. Saunders, Philadelphia (1986))

35

Hemoglobin is the blood protein which carries oxygen to tissues. It is present in large quantities in the

erythrocytes (red blood cells), which are little more than "bags" of hemoglobin. Hemoglobin messenger RNA (mRNA) is produced in pre-erythrocyte cells (erythroids). Because they lack a nucleus, erythrocytes are not capable of directing the manufacture of hemoglobin mRNA. Young erythrocytes in the blood, called reticulocytes, carry hemoglobin mRNA and translate it into protein.

Oxygen is picked up by hemoglobin in the lungs and released only in the oxygen-reduced, CO<sub>2</sub>-rich capillaries supplying the tissues. Hemoglobin carries four oxygen molecules at saturation (Po<sub>2</sub>>90 mm Hg), and releases only some of that oxygen in the deoxygenated state of the capillaries near the tissues and in the veins (Po<sub>2</sub>= 40 mm Hg). Other physiological conditions present near metabolizing tissues --high CO<sub>2</sub>, high acidity, high temperature, and high concentration of the compound 2,3-diphosphoglycerate (DPG)-- cause the release of more oxygen to the tissues at the venous Po<sub>2</sub> when it is demanded by high metabolism.

The hemoglobin protein consists of four amino-acid chains or subunits. The predominant hemoglobin in normal adults (about 92%) is called hemoglobin-A (HbA) and has two each of so-called  $\alpha$  chains and  $\beta$  chains. Its four-chain structure is denoted as  $\alpha_2\beta_2$ . A minor hemoglobin in normal adults, HbA<sub>2</sub>, consists of two  $\alpha$  chains and two  $\delta$  chains ( $\alpha_2\delta_2$ ). Other hemoglobin molecules are present in small amounts in normal adults.

Hemoglobin is highly-tuned by evolution to deliver oxygen as needed. For example, each stage of human development from embryo to adult has different oxygen needs and hemoglobins, which differ slightly in amino acid sequence. In early embryo, there are Gower-1 ( $\zeta_2\varepsilon_2$ ), Gower-2 ( $\alpha_2\varepsilon_2$ ) and Portland ( $\zeta_2\gamma_2$ ) hemoglobins. In the fetus and persisting for 5 to 6 months after birth, the predominant

hemoglobin is fetal hemoglobin (HbF;  $\alpha_2\gamma_2$ ). These normal hemoglobins and their amounts in different life phases are summarized in Bunn and Forget, "Hemoglobin: Molecular, Genetic and Clinical Aspects", W.B. Saunders, Philadelphia pp. 62, 68 (1986).

Defects in either the hemoglobin genes, the DNA sequences regulating transcription of the genes, or sequences involved in processing the messenger RNA (mRNA) can cause severe life-threatening and life-long illnesses.

Sickle cell anemia,  $\beta$ -thalassemia and  $\alpha$ -thalassemia denote the three major disease categories. Sickle cell anemia (SCA) and the much milder sickle cell trait are caused by a single amino acid change in the  $\beta$  chain. The defective  $\beta$  chain is denoted by  $\beta^S$  and hemoglobin containing  $\beta^S$  chains is denoted as HbS. The specific defect is a valine replacing the normal glutamic acid (Glu ---> Val), and the underlying DNA base mutation is adenine to thymine (A--->T). If the individual has inherited two defective genes (sickle cell anemia), all the  $\beta$ -containing hemoglobins have the structure  $\alpha_2\beta^S_2$ . If only one defective gene is inherited (sickle cell trait), the resulting hemoglobins are mixed:  $\alpha_2\beta^S_2$ ,  $\alpha_2\beta^S\beta$ , and  $\alpha_2\beta_2$ , so some normal hemoglobin is present. Relevant to the present invention is the fact that only some normal hemoglobin is sufficient to render SCA relatively harmless.

$\beta$ -thalassemia results from the absence of functional  $\beta$  chain. Over 100 different mutations have been associated with  $\beta$ -thalassemia (Collected in Huisman, *Hemoglobin*, 17:479 (1993)). Failure to produce functional  $\beta$  chain can result from, for example, nonsense and frameshift mutations in the gene itself, mutations in the regions that control gene transcription, or intervening sequence (IVS or intron) mutations that interfere with proper splicing of the mRNA.  $\beta$  thalassemia is classified as either  $\beta^+$  or  $\beta^0$ , depending on whether one or both  $\beta$  chain DNA regions are defective.

In the severe disease,  $\beta^0$  thalassemia, both are defective, and no  $\beta$  chains are produced. In the mild  $\beta^+$  disease, some, but very little  $\beta$  chain is produced.

In one embodiment of the TDR method of the present invention, a single therapeutic strategy of targeting a single third-strand to a region near the locations of different  $\beta$ -thalassemia mutations in different patients to allow homologous recombination with a single "normal sequence" donor strand can correct the several different mutations in the different patients. Thus, a single composition of matter (third strand and donor DNA) will provide therapy for patients with different underlying causes of disease.

$\beta^0$ -thalassemia is often much less severe if it is associated with elevation of other hemoglobins. For example, in a condition known as hereditary persistence of fetal hemoglobin (HPFH) in which fetal hemoglobin is synthesized into adult life, the presence of 20% to 30% HbF results in only mild disease (Bunn and Forget, *op. cit.*, pp. 345-346; Apperley, *Bailliere's Clinical Haematology*, 6:299 (1993)), and lesser amounts of HbF will reduce the severity of the disease (Perrine, *et al.*, *N. Engl. J. Med.* 328: 81 (1993)).

The severity of SCA is reduced by the presence of HbF as well. To eliminate most disease symptoms, it is estimated that 20% HbF is needed, and as little as 10% to 12% HbF can reduce or make infrequent some disease symptoms, such as protection against stroke (Noguchi, *et al.*, *N Engl. J. Med.* 318:96 (1988); Charache, *Experientia*. 49:126 (1993); Jackson, *et al.*, *J. Am. Med Assoc.* 177: 867 (1961); Davies, *Blood Reviews*. 7:4 (1993)).

#### Clinical Manifestations of Hemoglobinopathies.

For the heterozygous conditions where only one copy of a defective  $\beta$  region is carried, the resulting diseases, sickle cell trait and  $\beta^+$  thalassemia, are usually asymptomatic with mild anemia present in most individuals.



For sickle cell anemia, there is high risk of septicemia in infancy and early childhood; and in the more severe cases, there is high risk of childhood stroke. In addition, extreme anemia due to destruction of red blood cells and painful crises due to the blocking of capillaries by sickled cells (vaso-occlusion) are major manifestations of SCA. Growth and development abnormalities, damaged organs, and a number of other complications occur (Bunn and Forget, *op. cit.*, pp. 510-533). The disease course is variable with some patients following a severe course beginning shortly after birth with early childhood stroke, whereas other patients are infrequently ill (Powars and Hiti, *AJDC*, 147:1197 (1993)). About 30% of patients experience devastating disease with recurrent pain and vaso-occlusion crises that result in repeated strokes, chronic renal failure, etc. About 60% of patients have less severe disease, and 10% remain nearly symptom-free throughout life (reviewed in Apperley, *Balliere's Clinical Haematology*. 6: 299 (1993)).

For the 30% with severe disease, risky therapies such as bone marrow transplants are warranted; but because bone marrow transplants require an immunologically-matched bone-marrow donor and because of other clinical considerations, it is estimated at present that only 10% of SCA patients are candidates for bone marrow transplant (Davies, *Blood Reviews*. 7:4 (1993)). The risk of death from bone marrow transplants is about 10% (Davies, *Blood Reviews*. 7:4 (1993)), so they cannot be undertaken lightly. With the autologous bone marrow transplants (ABMT) embodied in this invention, the patient's own treated bone marrow is transplanted; and with other risk reduction factors, 30% of SCA patients would be candidates for bone marrow transplants.

In contrast to the variable clinical course in SCA, the clinical course for most cases of  $\beta$ -thalassemia is severe. Within a year after birth, severe hemolytic anemia develops, and regular transfusions are necessary to

maintain adequate hemoglobin levels. Most children on a regular transfusion program will develop normally and will have a good quality of life until cardiac complications due to iron overload develop in the mid teens to early twenties. Through damage to a number of organs including the heart and liver, iron overload is the major cause of morbidity and mortality in the second decade of life (Wayne, et al., *Blood*. 81: 1109 (1993); reviewed in Apperley, *Balliere's Clinical Haematology*. 6: 299 (1993); Bunn and Forget, *op. cit.*, pp. 351-356). Transfusions also carry the risk of complications and of diseases such as hepatitis and AIDS. Without adequate transfusion, morbidity and mortality occur sooner in life (Bunn and Forget, *op. cit.*, pp. 335-336).

Since  $\beta^0$ -thalassemia has such a severe clinical course, bone marrow transplants and other drastic therapies are justified despite the high risk of complications and even death. In the context of this invention, ABMT is less risky than current therapies.

While  $\alpha$ -thalassemias (caused by defective or missing  $\alpha$  chains) are known, and the methods, target sites, and third-strands directed to those sites, etc. apply to these conditions and other hemoglobinopathies as well, SCA and  $\beta^0$  thalassemia are more prevalent and will be used to illustrate the utility of the invention.

#### Molecular Biology and Molecular Physiology of Sickle Cell Anemia and $\beta$ -Thalassemia.

Clinical management of hemoglobinopathies, especially newer approaches, take advantage of a molecular-level understanding of these diseases. To anchor the subsequent description of current clinical management, new approaches to management, and the present invention, we summarize here the relevant molecular properties of normal and abnormal hemoglobins, their genes and gene control, and the relationship to disease.

# 1. Sickle cell anemia and hemoglobin-S polymerization

A single mutation in the  $\beta$  chain of hemoglobin is sufficient to cause the sickle-shaped cells characteristic of SCA. At the DNA level, the normal adenine base is mutated to a thymine, which causes an amino acid change from a negatively-charged glutamic acid to an uncharged, hydrophobic valine. The normal DNA sequence surrounding and including the changed base (boldface) is:

gag

5'...ATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTG...3' (SEQ ID NO:1)

3'...TACCACGTGGACTGAGGACTCCTCTTCAGACGGCAATGACGGGAC...5'

The altered DNA sequence surrounding and including the changed base (boldface) is:

gug

5'...ATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTG...3' (SEQ ID NO:2)

3'...TACCACGTGGACTGAGGACACCTCTTCAGACGGCAATGACGGGAC...5'

The codons "gag" and "gug" code for glutamic acid and valine, respectively.

In the oxygen-saturated state, the properties of sickle cell hemoglobin, HbS, in erythrocytes are essentially the same as the normal adult HbA; however, even the loss of one oxygen from HbS causes it to aggregate with other HbS and HbA molecules. High-molecular-weight, physically-large polymers consisting of 14 double-stranded fibers of hemoglobin molecules form, which in turn distort the erythrocytes into a characteristic sickle shape. The distorted erythrocytes cannot pass through the capillaries (vaso-occlusion) which is the cause of many of the severe medical problems associated with SCA, which include insufficient blood supply to tissues and organs with

subsequent damage, stroke, and sickle-cell pain crises. In addition, the distorted erythrocytes are selectively destroyed causing severe anemia. Prevention of aggregation, polymerization and subsequent sickling, then,  
5 is one way to manage the disease.

2. Regulation and processing of hemoglobin gene sequences in health and disease.

10 a. Normal  $\beta$ -chain gene regulation

Since erythrocytes contain no DNA, hemoglobin mRNA is synthesized in precursor erythroid cells. There are a number of factors that influence globin gene expression in erythroid cells (recently reviewed in: *Current Opinion in*  
15 *Genetics and Development*, 3: 232 (1993)). In overview, some aspects of gene regulation relevant to the present invention include:

Several DNA sequences are involved in regulating the level of mRNA transcripts for  $\beta$  chain synthesis. At or  
20 near the  $\beta$ -chain gene are regulatory elements in three separate locations: the promoter directly upstream from the gene, a sequence in the second intron, and a 3' flanking sequence (Behringer, et al., *Proc. Natl. Acad. Sci. USA*. 84: 7056 (1987); Killoias, et al., *Nucleic Acids Res.*, 15: 5739 (1987); Trudel, et al., *Mol. Cell. Biol.* 7: 4024  
25 (1987); Antoniou, *EMBO J.* 7: 377 (1988); deBoer, et al., *EMBO J.* 7: 4203 (1988)). In addition, there are sequences somewhat distant from the  $\beta$ -chain gene that help control the level of transcription and tissue specificity. Two  
30 such sequences are the locus control region (LCR) and an enhancer element (ENH). These distant sequences may interact in ways, which are not yet understood, to adjust levels of both  $\beta$ - and  $\gamma$ -chain synthesis simultaneously (Balta, et al., *Blood*. 83: 3727 (1994)). Finally, trans-  
35 acting factors (e.g., proteins or other molecules) may increase or decrease hemoglobin gene expression (Bunn and Forget, op. cit., pp. 192-197).

Relevant to this invention is that it is unlikely that standard gene therapy methods, which employ viral vectors that either integrate at non-native sites in the genome or do not integrate at all, can reproduce this complicated normal control of hemoglobin gene expression. In contrast, the TDR method of the present invention replaces the DNA at the native site with the donor DNA, so regulation of gene expression will follow the native course, provided that the donor DNA was not purposely designed to alter the native course of gene expression.

b.  $\beta$ -chain gene regulation in  $\beta$ -thalassemia

$\beta^0$ -thalassemia can, in principle, be caused by any  $\beta$ -gene-associated mutation that completely deactivates HbA. It is not surprising, then, that  $\beta^0$ -thalassemia is caused by kinds of mutations that include nonsense and frameshift mutations and mutations that cause defective processing of the mRNA. For many specific mutations the exact base change is known. A summary and discussion of specific base mutations, as of 1988, may be found in Bunn and Forget (*op. cit.*, p. 274). A recent summary of all known mutations may be found in Huisman (*Hemoglobin*, 17: 479 (1993)). Relevant to this invention is the fact that a single or a few donor DNAs in the TDR method can be used to correct all approximately 100 known  $\beta$ -thalassemia mutations.

Other thalassemias are caused by improper processing of introns (intervening sequences or IVS). Of particular interest, at the 3' ends of introns the consensus sequence for optimal mRNA splicing is:

(T or C)<sub>n</sub>N(C or T)AGG-3'

where  $n > 10$ , N stands for any nucleotide, and the underlined AG sequence is invariant among many species, and therefore, is thought to be absolutely required for proper splicing (Bunn and Forget, *op. cit.*, pp. 177-178). Relevant to this invention, the presence of the (T or C)<sub>n</sub> sequence provides

a third-strand polypurine target site on the opposite strand.

5 Some Newer Clinical Approaches to Management and Cure of SCA and  $\beta$ -Thalassemia.

Because of the long-term complications of and risks associated with blood transfusions and because of the mortality and high failure rate of bone marrow transplants in the treatment of SCA and  $\beta$  thalassemia, there is a dire  
10 need for better and safer treatments for these diseases. Since autologous bone marrow transplants are a preferred protocol of the invention, issues and procedures in bone marrow transplants will be briefly discussed here.

15 1. Bone marrow transplants

Bone marrow transplants are used to treat both sickle-cell anemia and  $\beta^0$  thalassemia. The general procedure is as follows:

Immune suppressants such as cyclophosphamide, immuran  
20 and azothiopain are administered to the patient to destroy some of the bone marrow to decrease substantially the risk of transplant immune rejection (graft vs. host disease or GVHD). In addition, immune depressants serve to decrease the numbers of abnormal embryonic stem cells in the SCA  
25 patient, which is desired to yield a high percentage of healthy, transplanted cells from the donated bone marrow. In  $\beta^0$  thalassemia, however, the affected pre-erythrocytes produce no functional hemoglobin, so their numbers need not be suppressed in advance of transplantation. Immune  
30 depressants leave the patient vulnerable to disease, which is a major reason for the 10% mortality rate associated with them.

Bone marrow from matched siblings, who are disease-free, is then injected intravenously to "home" to the bone  
35 marrow of the patient. Siblings who have been matched for immune-compatibility are used; otherwise, transplant rejection and GVHD are too frequent a complication. A

major problem is that many SCA and  $\beta^0$ -thalassemia patients do not have a disease-free, matched sibling to serve as bone marrow donor, which severely limits the applicability of bone marrow transplants.

5 Since introduction of bone-marrow from another person is problematic, even if the donor and patient are closely matched immunologically, clinical research in bone marrow transplants is moving in the direction of treating the patient's own bone marrow and then returning it to the body  
10 (autologous bone marrow transplant, or ABMT). The methods of this invention are consistent with this important direction of bone marrow transplant research and clinical application.

15 Until very recently, hematopoietic stem cells (that are the targets for the therapies of this invention and in general for gene therapies for these diseases in general) could be proliferated ex vivo for only a few generations. Berardi, et al. (Science, 267: 104. (1995)), have just developed a stem cell isolation procedure that provides  
20 primitive hematopoietic cells in high concentrations (1 in  $10^5$  of bone marrow mononuclear cells), so this area of stem-cell therapies is advancing rapidly. In addition, the isolated cells proliferate both along lymphoid and myeloid (precursors to erythrocytes) lineages, and can be made to  
25 form sizable colonies and secondary cultures on replating with an efficiency of 40%. These numbers are encouraging for providing for replanting into a patient sufficient numbers of hematopoietic stem cells altered by the methods of this invention.

30 The present invention can circumvent a number of the problems with present-day bone-marrow-transplant technology. Since the transplanted bone-marrow is treated bone marrow obtained from the patient, immune rejection and GVHD should be reduced or eliminated; thus eliminating the  
35 need for immune suppressants at least for immune-rejection reasons. For SCA, immune depressants may still be required to destroy some of the  $\beta^s$  producing cells before

transplantation. For  $\beta^0$ -thalassemia, however, immune depressants might be eliminated entirely since the diseased cells may not need to be destroyed. The invention should then reduce or eliminate much of the morbidity and mortality associated with bone marrow transplants.

#### SUMMARY OF THE INVENTION

The present invention provides a method for effecting repair of genetic defects in the  $\beta$  globin gene and DNA regions involved in the expression of that gene. In one method (the TDR method), repair is effected through homologous recombination between a native nucleic acid segment in the  $\beta$ -gene cluster on chromosome 11 in a human cell and a donor nucleic acid segment introduced into the cell, which comprises:

- a) introducing into a human cell i) an oligonucleotide third strand which comprises a base sequence capable of forming a triple helix at a binding region on one or both strands of a native nucleic acid segment, said native nucleic acid segment containing an undesired mutation in the vicinity of the human  $\beta$  globin gene cluster target region where the recombination is to occur, said oligonucleotide being capable of inducing homologous recombination at the target region of the native nucleic acid, and ii) a donor nucleic acid which comprises a nucleic acid sequence sufficiently homologous to the native nucleic acid segment such that the donor sequence is capable of undergoing homologous recombination with the native sequence at the target region;
- b) allowing the oligonucleotide third strand to bind to the native nucleic acid segment to form a triple stranded nucleic acid, thereby inducing homologous recombination at the native nucleic acid segment target region; and
- c) allowing homologous recombination to occur between the native and donor nucleic acid segments.

Another aspect of the present invention concerns a



method for effecting homologous recombination between a native nucleic acid segment in a cell and a donor nucleic acid segment introduced into the cell, which comprises:

- 5 a) contacting a donor nucleic acid segment with an oligonucleotide third strand which comprises a base sequence capable of forming a triple helix at a binding region on one or both strands of the donor nucleic acid segment in the vicinity of a target region where the recombination is to occur, said oligonucleotide being  
10 capable of inducing homologous recombination at the target region of the donor nucleic acid, and said donor having a sequence sufficiently homologous to a first nucleic acid segment in a human cell which comprises at least a portion of the human  $\beta$  globin gene cluster, such that the donor  
15 sequence will undergo homologous recombination with the first sequence at the target region;
- b) treating the nucleic acid segment by allowing the oligonucleotide to bind to the donor nucleic acid segment to form a triple stranded nucleic acid;
- 20 c) introducing into a human cell the treated donor nucleic acid; and
- d) allowing homologous recombination to occur between the first and donor nucleic acid segments.

Another aspect of the present invention concerns a  
25 method for causing a mutation at a specific DNA sequence site in a cell, which comprises:

- a) introducing into a cell an oligonucleotide third strand which comprises a base sequence which is capable of forming a triple helix at a binding region on one or both  
30 strands of a native nucleic acid segment which contains an undesired mutation in the vicinity of the human  $\beta$  globin gene cluster target region, said oligonucleotide being capable of inducing mutagenesis when bound to the binding region;
- 35 b) allowing the oligonucleotide to bind to the native nucleic acid segment to form a triple stranded nucleic acid; and

c) allowing mutagenesis to occur in the target region.

In another aspect, the present invention provides a composition comprising:

5 a) an oligonucleotide third strand which comprises a base sequence which is capable of forming a triple helix at a binding region on one or both strands of a native nucleic acid segment in the vicinity of the human  $\beta$  globin gene cluster target region, said oligonucleotide being capable  
10 of inducing homologous recombination at the target region of the native nucleic acid; and

b) a donor nucleic acid which comprises a nucleic acid sequence sufficiently homologous to the native nucleic acid segment such that the donor nucleic acid will undergo  
15 homologous recombination with the native sequence at the target region when the third strand is bound to the native nucleic acid.

In another aspect, the present invention provides a composition comprising an oligonucleotide third strand  
20 which comprises a base sequence capable of third strand binding to a portion of one or both strands of a native nucleic acid segment in the human  $\beta$  globin gene-cluster of a cell, said oligonucleotide being capable of causing a mutation at a specific site of the native nucleic acid  
25 segment when bound to the native nucleic acid segment.

The present methods and compositions are useful in research and therapeutic applications where site specific recombination in the human  $\beta$  globin gene-cluster of a cell is desired. The present inventions are also useful  
30 constructing cell lines and transgenic animals for the study of hemoglobinopathies.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic illustration of the targeted  
35 mutagenesis (TM) method using a psoralen-linked third strand as an example. The double stranded Watson-Crick binding is indicated by "=", while third strand binding is

indicated by "\*". The AT base pair in boldface is the base pair to be changed to a "normal" TA base pair. Long wavelength UV light (UVA) photoactivates the psoralen generating a psoralen adduct on the T in the AT base pair.

5 The damage to the AT base pair is cross-linking as indicated by "+". The damage is initiated through the psoralen reacting specifically with the T base in the AT pair, and the other base pairs in the DNA remain unchanged and are not shown in the center and right-hand parts of the  
10 Figure. The cell's native DNA repair machinery attempts to repair the damage, but the machinery often makes a mistake and repairs the cross link to a TA base pair instead of the original AT base pair.

Figure 2 is an illustration of the targeted DNA  
15 replacement method. Figure 2(A) is a schematic illustration of a mutation responsible for a genetic defect, x, and a third-strand binding site downstream from the mutation. The genetic defect may be a base substitution, deletion or addition of one or more bases.  
20 The third strand binding site may be located thousands of bases from the defect. In Figure 2(B), the mutagenic third strand binds to its targeted purine site and causes DNA damage. In Figure 2(C), the native cellular machinery, stimulated by the DNA damage, "aligns" the donor DNA with  
25 the defective chromosome region to allow homologous recombination to occur between the donor double strand and the chromosome defective region, which results in a repaired chromosome region, shown in Figure 2(D).

Figure 3 illustrates the approximate relative  
30 locations of the  $\beta$ -globin and  $\beta$ -globin-like genes on the human chromosome 11  $\beta$ -gene cluster. The cluster spans about 52 kilobases from the beginning of the embryonic  $\epsilon$  gene through the adult  $\beta$  gene. The  $\alpha$  gene cluster is also shown. (Reproduced from Bunn and Forget, "Hemoglobin:  
35 Molecular, Genetic and Clinical Aspects", W.B. Saunders, Philadelphia (1986), p. 174).

DETAILED DESCRIPTION OF THE INVENTIONThird Strand Binding

5 For purposes of illustrating the present invention, five binding motifs are described. It is understood that practice of the invention is not limited to these motifs. Table 1 summarizes these five motifs, which are additional rules subject to the binding code. The motifs provide  
 10 further instructions for defining the sequences of different third-strands that will specifically and stably bind to a single purine center-strand target. The Table also shows selected analog bases which may be substituted for the native A, G, T, and C third-strand bases.

Table 1

20	Motif Description: Third-Strand Bases/Strand Polarity	Binding Code	Some Analog Substitutions
	Pyrimidine/parallel	T:AT C+:GC	me <sup>5</sup> C <sup>+</sup> for C <sup>+</sup> propyne <sup>5</sup> C <sup>+</sup> for C <sup>+</sup> propyne <sup>5</sup> U for T
	Purine/parallel (A-rich targets)	A:AT G:GC	2,6 DAP for A
25	Purine/antiparallel (G-rich targets)	A:AT G:GC	2,6 DAP for A
	T and G/parallel (high nearest neighbor frequencies for AA, GG in center strand)	T:AT G:GC	7-deaza-2'- deoxyxanthosine for T
30	T and G/antiparallel (high nearest neighbor frequencies for AG, GA in center strand)	T:AT G:GC	propyne <sup>5</sup> U for T

In the Binding Code column, the colon indicates third-strand binding of the base to the left of the colon to the  
 35 center base to the immediate right of the colon. The +

superscript indicates that the bases are in the protonated form when they bind (the energy of binding provides the energy for protonation). DAP stands for di-aminopurine.

To form a stable triplex, third-strands should  
5 preferably be at least about 10 bases in length, more preferably at least about 20 bases. The probabilities of target purine runs of these lengths at any one site in a random genome are  $1/2^{10}=9.8 \times 10^{-4}$  and  $1/2^{20}=9.5 \times 10^{-7}$ , respectively. There are a number of general approaches for  
10 increasing available targets:

- Widening the binding code to include pyrimidine bases in the center strand;

- Designing third strands that bind to a purine run in  
15 one strand and then can bind to an adjacent purine run in the other strand of the Watson-Crick helix (here called strand switching or alternative third-strand recognition); and

- Providing bases in the third strand opposite  
20 mismatches (i.e., occasional pyrimidines in the center strand) that are more energetically favorable than other bases.

Conversely, there are a number of approaches that are aimed  
25 at reducing the length of purine runs that may be targeted:

- Designing analog bases with higher binding affinity;
- Designing analog backbones with lower binding  
enthalpy or that experience less entropy decrease upon  
30 binding; and

- Incorporating in the third strand molecules which intercalate (e.g., acridine) or otherwise favorably interact with the double helix.

35 These approaches are reviewed in (Sun and Helene, *Current Opinion in Structural Biology*. 3: 345 (1993)). It will be understood that the scope of the practice of this

invention includes all the approaches above. In one embodiment of the invention, the targeted DNA replacement method (see below), the third-strand binding site may be thousands of bases from the site of the genetic defect, so there is a high probability that at least one purine run of sufficient length will be found. For example, there are on average at least  $9.8 \times 10^{-4} \times 10,000 = 9.8$  purine runs of ten bases within 10,000 bases of the site of a genetic defect.

Examples of analog bases for various motifs include, but are not limited to, those presented in Table 1. Other analog bases, for example, are discussed in Sun and Helene (op. cit.) and also in co-pending U.S. patent application entitled "Residues for Binding Third Strands to Complementary Nucleic Acid Duplexes of any Base-Pair Sequence", filed concurrently herewith, the content of which is incorporated by reference. Fresco and coworkers (Fossella, et al., *Nucleic Acids Res.* 21: 4511 (1993)) have identified bases opposite pyrimidines in the center strand (mismatches) in the pyrimidine/parallel motif of Table 1 that only minimally destabilize the triple helix. The relative stability of these mismatches as measured by melting temperature of the triple helix in which they are present are presented in Table 2, which shows the effect of third-strand bases in the center of a 21-mer triple helix in the pyrimidine/parallel motif as measured by melting temperature of the third-strand. The test helices were composed of the single strands  $A_{10}-X-A_{10}$  (Watson-Crick center strand),  $T_{10}-Y-T_{10}$  (other Watson-Crick strand),  $T_{10}-Z-T_{10}$  (third strand). The triple-helix bases at the mid position in the helix is denoted by Z:XY. The body of the table is melting temperature in °C. Choice of the most stable base at pyrimidine center-strand sites is useful for designing third strands to purine runs interrupted by occasional pyrimidines. The table shows, for example in the pyrimidine/parallel motif, that a G base in the third strand opposite a T base mismatch in the center strand leads to a more stable third strand, as its melting

temperature is 16.1°C vs. -5.0°C, -0.3°C, and -3.0°C for the A, T and C bases.

5

Table 2

	Watson Crick Base Pair			
Base (Z) in Third Strand	AT	TA	GC	CG
A	3.0	-5.0	13.5	3.2
G	4.3	16.1	9.6	7.8
T	22.6	-0.3	18.3	14.0
C	3.8	-3.0	31.0	10.0

10

15

20

25

30

35

The existence of both parallel and antiparallel motifs can be used simply for strand switching or alternative strand recognition. Since the two Watson-Crick strands are antiparallel, when a third strand with a normal 5'-3' backbone switches from a purine run on one strand to a purine run on the other strand, its binding to the center strand will switch from parallel to antiparallel, or vice versa. Or alternatively, if parallel or antiparallel binding is to be preserved after the switch a 3'-3' or 5'-5' linker must be provided at the switch point. Such linkers are well-known to those skilled in the art and are commercially available. In addition, the linkers must provide enough flexibility so that third-strand binding is not sterically hindered by strand switching. The use of strand switching in third-strand binding is well documented in the literature (U.S. Patent No. 5,399,676; Horne and Dervan, *J. Amer. Chem. Soc.*, 112, 2435 (1990); Jayasena and Johnston, *Nucleic Acids Res.* 20: 5279 (1992); Sumedha and Johnston, *Biochemistry.* 32: 2800 (1993); Froehler, et al., *Biochemistry.* 31: 1603 (1992)). Relevant to the present invention, for example, the  $\beta$  globin gene site where the

mutation responsible for sickle cell anemia is found has three adjacent purine runs switching from strand to strand (see below and Example 1).

5     Third Strand-Targeted Mutagenesis (the TM method)

          This method for genetic defect repair takes advantage of misrepair --that is, mistakes made by the cellular mechanisms involved in the repair of mutagen-damaged DNA. Using psoralen-linked third-strands as an example, genetic defect repair is accomplished as illustrated in Figure 1 and is briefly described below. See also copending U.S. application 08/083,088 filed June 25, 1993, and published as WO 95/01364 on January 12, 1995, entitled "Chemically Modified Oligonucleotide for Site-Directed Mutagenesis", the contents of which are incorporated by reference.

          A third-strand targeted to the site of the genetic defect is prepared with a mutagen, preferably psoralen, attached to its end. Psoralen selectively reacts with the base T. The psoralen-linked third strand is then introduced into cells in culture removed from the patient (ex vivo therapy). Alternatively, the mutagen-linked third strands may be injected or delivered by intravenous infusion. The third-strand binds specifically to a double-stranded chromosomal DNA sequence according to the third-strand binding code. The cell culture is then bathed in long-wavelength ultraviolet light (centered at 365 nm, also known as UVA), which causes the psoralen to damage the double-stranded DNA target by cross-linking the two strands together at the AT base-pair target site (boldface type). The cell's native DNA repair mechanism recognizes the damage, and attempts to repair the damaged T base, but frequently makes the mistake of replacing the T base with an A base.

35           Some characteristics of the repair process are:

          -A single dose of psoralen-third-strand drug induces



genetic defect repair in over 6% of cells in one animal cell culture system. Genetic diseases that require for amelioration or cure less than 20% repair will likely need only 1 to 4 doses of drug.

5       -For the mutagen psoralen, the cellular-repair-process mistake almost always changes a T base to an A base. Other mutagens can preferentially target different bases and misrepair will result in different bases.

10       -The targeted T to A base change is about 100-times more prevalent than background mutations caused by free psoralen (psoralen not linked to the third strand). Thus, third-strands promote damage and misrepair with high specificity.

15       Third-strand-targeted homologous recombination or targeted DNA replacement (the TDR method)

      This method for genetic defect repair involves targeting third strands to native DNA sequences to induce DNA damage to stimulate homologous recombination with an  
20       introduced donor nucleic acid strand. In another embodiment, the third strand modifies or damages the donor nucleic acid before it is introduced into the target cell. In particular, the invention provides a method for effecting gene transfer or mutation repair at a specific  
25       sequence site on the target native nucleic acid in a cell. The method utilizes two nucleic acids: (1) an oligonucleotide third strand capable of specifically binding to the binding region of a native double-stranded nucleic acid, and (2) a donor nucleic acid fragment capable  
30       of undergoing homologous recombination with the native nucleic acid targeted by the oligonucleotide. The nucleic acid sequence of the donor DNA is slightly different from the native nucleic acid it is replacing by homologous recombination to, for example, repair a genetic defect.  
35       The method and some of its features are illustrated in Figure 2.

      This method, referred to as the TDR method, is more

general than the TM method for two reasons: First, the genetic defect to be repaired is not required to be near a third strand binding site. In fact, the genetic defect may be thousands of bases distant from the third strand binding site. Thus, most genetic defects are potential therapeutic targets, compared to the TM method where therapeutic targets must be selected to be in or near third-strand binding sites. Secondly, the TDR method is able to correct multiple base substitutions or small or large base deletions and/or insertions, as long as the donor nucleic acid has acceptable homology with the native nucleic acid it is replacing. In targeted mutagenesis, in contrast, the genetic defect to be repaired is usually restricted to a single base substitution, or occasionally to a single base deletion or addition.

Targeted DNA replacement accomplishes the same therapeutic goals as viral-vector gene therapies, but with a number of advantages:

-The repaired gene resides at its native chromosomal site, so repair and hence disease cure is permanent, and gene expression will be native. In contrast, viral vectors either provide only temporary cure or they integrate into chromosomes in non-native sites, so patterns of gene expression may not be compatible with the requirements of a disease cure.

-Donor nucleic acids smaller than whole genes may be used, which may be delivered by standard methods, IV or injection.

-Most human cell types are available for targeting. In contrast, most viral vectors are targeted to a very limited number of cell types.

#### Oligonucleotide

The oligonucleotide third strand useful in either the TM or TDR methods is a synthetic or isolated oligonucleotide capable of binding with specificity to a

predetermined binding region of a double-stranded native nucleic acid molecule to form a triple-stranded structure. The third strand may bind solely to one strand of the native nucleic acid molecule, or may bind to both strands at different points along its length. In the practice of the TDR method, the predetermined target region of the double-stranded DNA is in or adjacent to a gene, mRNA synthesis or processing control region, or other DNA region that it is desirable to replace by homologous recombination. The predetermined binding region, if adjacent to the targeted region, is preferably within 10,000 nucleotides or bases from the targeted region.

Preferably, the oligonucleotide is a single-stranded DNA molecule between about 7 and about 50, most preferably between about 10 and about 30 nucleotides in length. The base composition can be homopurine, homopyrimidine, or mixtures of the two. The third strand binding code and preferred conditions under which a triple-stranded helix will form are well known to those skilled in the art (Fresco U.S. Patent 5,422,251; Beal and Dervan, *Science* 251: 1360 (1991); Beal and Dervan, *Nucleic Acids Res.*, 20:2773 (1992); Broitman and Fresco, *Proc. Natl. Acad. Sci. USA*, 84:5120 (1987); Fossella, et al., *Nuc. Acids Res.* 21:4511 (1993); Letai, et al., *Biochemistry* 27:9108 (1988); Sun, et al., *Proc. Natl. Acad. Sci. USA* 86:9198 (1989)), and are described above. The third strand need not be perfectly complementary to the duplex, but may be substantially complementary. In general, by substantially complementary is meant that one mismatch is tolerable in every about 10 base pairs.

The oligonucleotide may have a native phosphodiester backbone or may be comprised of other backbone chemical groups or mixtures of chemical groups which do not prevent the triple-stranded helix from forming. These alternative chemical groups include phosphorothioates, methylphosphonates, peptide nucleic acids (PNAs), and others known to those skilled in the art. Preferably, the oligonucleotide

backbone is phosphodiester.

The oligonucleotide may also comprise one or more modified sugars, which would be well known to one of ordinary skill. An example of such a sugar includes  $\alpha$ -  
5 enantiomers.

The third strand may also incorporate one or more synthetic bases if such is necessary or desirable to improve third strand binding. Examples of synthetic base design and the bases so designed are found in the co-  
10 pending U.S. application of Fresco, et al. entitled "Residues for Binding Third Strands to Complementary Nucleic Acid Duplexes of any Base-Pair Sequence", filed concurrently herewith.

If it is desired to protect the oligonucleotide from  
15 nucleases resident in the target cells, the oligonucleotide may be modified with one or more protective groups. In a preferred embodiment, the 3' and 5' ends may be capped with a number of chemical groups such as an alkyl amine group, a thiol group, cholesterol, acridine, etc. In another  
20 embodiment, the oligonucleotide third strand may be protected from exonucleases by circularization.

The oligonucleotide third strand should be capable of inducing either homologous recombination or targeted mutagenesis at a target region of the native nucleic acid.  
25 That may be accomplished by the binding of the third strand alone to the native nucleic acid binding region, or by a moiety attached to the oligonucleotide. In the embodiment where the binding of the third strand alone induces the recombination, the third strand should bind tightly to the  
30 binding region, i.e., it should have a low dissociation constant ( $K_d$ ) for the binding region. The  $K_d$  is estimated as the concentration of oligonucleotide at which triplex formation is half-maximal. Preferably, the oligonucleotide has a  $K_d$  less than or equal to about  $10^{-7}$  M, most preferably  
35 less than or equal to about  $2 \times 10^{-8}$  M. The  $K_d$  may be readily determined by one of ordinary skill, including

estimation using a gel mobility shift assay (Durland, et al., *Biochemistry* 30, 9246 (1991); see also the copending U.S. application of Glazer entitled "Triple Helix Forming Oligonucleotides for Targeted Mutagenesis" filed  
5 concurrently herewith, the content of which is incorporated by reference.).

### Mutagen

The oligonucleotide may be chemically modified to  
10 include a mutagen at either the 5' end, 3' end, or internal portion so that the mutagen is proximal to a site where it will cause damage to the native nucleic acid. Preferably the mutagen is incorporated into the oligonucleotide during nucleotide synthesis. For example, commercially available  
15 compounds such as psoralen C2 phosphoroamidite (Glen Research, Sterling VA) are inserted into a specific location within an oligonucleotide sequence in accordance with the methods of Takasugi et al., *Proc. Natl. Acad. Sci. USA*, 88:5602 (1991); Gia et al., *Biochemistry* 31:11818  
20 (1992); Giovannangeli, et al., *Proc. Natl. Acad. Sci. USA*, 89:8631 (1992), all of which are incorporated by reference herein.

The mutagen may also be attached to the oligonucleotide through a linker, such as sulfo-m-  
25 maleimidobenzoyl-N-hydroxysuccinimide ester (sulfo-MBS, Pierce Chemical Company, Rockford IL) in accordance with the methods of Liu et al., *Biochem.* 18:690 (1979) and Kitagawa and Ailawa, *J. Biochem.* 79:233 (1976), both of which are incorporated by reference herein. Alternatively,  
30 the mutagen is attached to the oligonucleotide by photoactivation, which causes a mutagen, such as psoralen, to bind to the oligonucleotide.

The mutagen can be any chemical capable of stimulating either mutagenesis or homologous recombination. Such  
35 stimulation can be caused by modifying the native nucleic acid in some way, such as by damaging with, for example, crosslinkers or alkylating agents. The mutagen may also be

a moiety which increases the binding of the third strand to the target, such as intercalators (e.g., acridine). Such mutagens are well known to those skilled in the art. The chemical mutagen can either cause the mutation

5 spontaneously or subsequent to activation of the mutagen, such as, for example by exposure to light.

Preferred mutagens include psoralen and substituted psoralens such as hydroxymethyl-psoralen (HMT) that require activation by ultraviolet light; bleomycin, fullerenes, 10 mitomycin C, polycyclic aromatic carcinogens such as 1-nitropyrene, alkylating agents; restriction enzymes, nucleases, radionuclides such as  $^{125}\text{I}$ ,  $^{35}\text{S}$  and  $^{32}\text{P}$ ; and molecules that interact with radiation to become mutagenic, such as boron that interacts with neutron capture and 15 iodine that interacts with auger electrons.

If necessary for activation of the mutagen, light can be delivered to cells on the surface of the body, such as skin cells, by exposure of the area requiring treatment to a conventional light source. Light can be delivered to 20 cells within the body by fiber optics or laser by methods known to those skilled in the art. Targeted fluorogens that provide sufficient light to activate the mutagens can also provide a useful light source. Ex-vivo exposure to light of cells such as embryonic stem cells can be carried 25 out by procedures known to those skilled in the art of ex vivo medical treatments.

#### Donor Nucleic Acid

The donor nucleic acid used in the practice of the TDR 30 embodiment is either a double-stranded nucleic acid, a substantially complementary pair of single stranded nucleic acids, or a single stranded nucleic acid. The sequence of the donor nucleic acid at its ends is substantially homologous to the nucleic acid region which is to be 35 replaced by homologous recombination. Preferably, the region of substantial homology is at least about 20 bases at each end of the donor nucleic acid. By "substantial

homology" is meant that at least about 85% of the available base pairs are matching.

The differences in base sequences between the donor nucleic acid and the targeted region it is desired to replace are base changes, deletions of bases or insertions of bases, nucleotide repeats, or a combination of these, chosen to accomplish the desired genetic and phenotypic change. Nucleic acid segments may be added according to the present invention. Such segments include a gene, a part of a gene, a gene control region, an intron, a splice junction, a transposable element, a site specific recombination sequence, and combinations thereof.

The donor nucleic acid strands, whether single- or double-stranded, may be gene sized, or greater or smaller. Preferably, they are at least about 40 bases in length, preferably between about 40 and about 1,000,000 bases in length. Most preferably, the lengths are between about 500 and about 3,000 bases.

#### 20 Method of Extracting Bone Marrow from a Patient

Extraction of bone marrow from a patient to obtain hematopoietic stem cells or erythrocyte precursors for treatment outside the patient's body (ex vivo treatment) is well know to those skilled in the art. The standard procedures may be found in, for example, Hoffman, et al., Hematology, Churchill-Livingstone, New York, 1995.

#### Method of Obtaining Hematopoietic Stem Cells in Culture

In order to treat hematopoietic stem cells or erythrocyte precursor cells with the methods and compositions of matter of this invention, it is necessary to isolate, maintain and proliferate the cells in culture outside the body. A recently developed procedure (Berardi, et al., Science, 267:104 (1995)) that utilizes antimetabolites to kill non-primitive bone marrow cells provides sufficient stem cells for the present purposes. Alternate procedures, such as administration of colony

stimulating factor-granulocyte (CSF-G), can also be used to mobilize stem cells from the marrow into the peripheral blood. Peripheral blood hematopoietic precursor and stem cells can then be collected by leukapheresis and used for the present invention. Techniques for isolation of hematopoietic stem cells are well known to those skilled in the art, and may also be found in Hoffman, et al., *Hematology*, Churchill-Livingstone, New York, 1995.

10 Method of Administration of the Oligonucleotide

Experimental manipulations such as electroporation, micro-injection, microprojectiles, calcium phosphate or other treatments well known to those skilled in the art may be used to deliver the oligonucleotide to the nucleus of the target cell. Preferably, the oligonucleotide can be delivered to cells or live animals simply by exposing the cells to the oligonucleotide by including it in the medium surrounding the cells, or in live animals or humans by bolus injection or continuous infusion. The exact concentration will be readily determined by one of ordinary skill, and will depend on the specific pharmacology and pharmacokinetic situation. Typically, from about 0.1 to about 10  $\mu\text{M}$  will be sufficient.

25 Modifying Donor Nucleic Acid

In another aspect of the invention, the nucleic acid modification or damage used to stimulate homologous recombination is targeted to the donor nucleic acid (as opposed to the native nucleic acid) either inside or outside the target cells. For the preferred embodiment where the nucleic acid modification or damage is effected outside the target cell, the modified or damaged donor nucleic acid is then introduced into the target cell to stimulate homologous recombination with the native nucleic acid.

Modifying or damaging the donor nucleic acid outside the cell has several desirable features including: nucleic



acid modification or damage can be caused with higher efficiency outside the cell; mutagens and other treatments (e.g., psoralen-UVA) potentially toxic to the cell, animal or human can be used since the mutagen can be isolated away from the modified or damaged donor nucleic acid before the purified donor nucleic acid is introduced into the target cell; conditions (e.g., temperature, cation composition and concentration) can be controlled to maximize binding of third-strands for any binding motif; and nucleic acid modifying or damaging agents can be directly synthesized into specific sites on the donor nucleic acid by methods well known to those skilled in the art, without the use of third strands. In addition, to increase efficiency of and to control the location of modification or damage, third strand sites can be engineered into the donor nucleic acid at a location where the engineered nucleic acid segment is unlikely to cause unwanted effects when the donor nucleic acid is recombined into the organism's native nucleic acid. The preferred method of introducing the oligonucleotide into stem cells and erythrocyte precursors is co-incubation in the growth medium with or without the addition of cationic liposomes (Wang, et al., *Mol. and Cell Biol.*, 15:1759-1768 (1995)).

#### 25 Method of Administration of the Donor Nucleic Acid

When using the TDR embodiment of the present invention, the donor nucleic acid can be delivered to the nucleus of cells in culture or cells removed from an animal or a patient (ex vivo) by manipulations such as peptide-facilitated uptake, electroporation, calcium chloride, micro-injection, microprojectiles or other treatments well known to those skilled in the art. For single-stranded donor nucleic acids of less than 100 bases, the donor nucleic acid can be delivered to cells or live animals or humans simply by exposing the cells to the oligonucleotide that is included in the medium surrounding the cells, or in live animals by bolus injection or continuous infusion.

One of the complementary single strands is delivered and the other delivered at the same time or up to 12 hours later, preferably from about 20 to about 40 minutes later.

5 The donor nucleic acid may also be introduced into the cell in the form of a packaging system which would be well known to one of ordinary skill. Such systems include DNA viruses, RNA viruses, and liposomes as in traditional gene therapy.

10 The preferred method of introducing the donor DNA into stem cells and erythrocyte precursors is co-incubation in the growth medium with or without the addition of cationic liposomes (Wang, et al., *Mol. and Cell Biol.*, 15:1759-1768 (1995)).

#### 15 Auxiliary Patient Treatments

To increase the percentages of treated hematopoietic stem cells or erythrocyte precursor cells in bone marrow, after the bone marrow has been removed from the patient for treatment, the patient may undergo chemotherapy to reduce or destroy the bone marrow cells in the body. These treatments are well known to those skilled in the art of autologous bone marrow transplants, see Hoffman, et al., *Hematology*, Churchill-Livingstone, New York, 1995.

#### 25 Method of Reintroducing Treated Cells.

Treated stem cells can be reintroduced into the patient by intravenous infusion, using standard methods.

#### Method of Use

30 The invention provides a method for effecting gene transfer, genetic defect repair, and targeted mutagenesis at a specific sequence site on the DNA target in the  $\beta$  globin gene cluster in cells such as stem cells and erythrocyte precursor cells of humans or other animals.

35 Examples of therapeutic use are apparent. For example, if a targeted DNA region contains base changes, deletions or additions of bases which cause an inherited or

somatic hemoglobinopathy such as sickle cell anemia, or  $\alpha$ -thalassemia or  $\beta$ -thalassemia, then the donor nucleic acid can provide a normal gene by replacing the defective nucleic acid to correct that disorder. If the inherited or  
5 somatic hemoglobinopathy is caused by a single base substitution (or certain specific single base deletions and additions) at a third-strand binding site, the hemoglobinopathy may be treated by an oligonucleotide carrying a mutagen, the modification or damage from which  
10 is subsequently misrepaired to provide an active, normal or near-normal hemoglobin.

The preferred embodiment of the methods and compositions of the invention are in ex vivo therapies where third strands and donor DNA can be introduced into  
15 target cells outside the body. The hemoglobinopathies are particularly amenable to ex vivo treatments.

The  $\beta$ -globin gene cluster (Figure 3) is approximately 52 kilobases (kb) in length, and a large number of purine runs of sufficient length (10 bases or greater) and a  
20 number of purine runs of the preferred minimum length (greater than 20 bases) are present in the cluster (see the Examples). Utilizing a donor DNA of length greater than 40 kb, the targeted DNA replacement method of this invention can carry out homologous recombination initiated by DNA  
25 damage at 40,000 bases or more from the site of the genetic defect to be corrected. Therefore, all genetic defects within the  $\beta$  gene cluster may be corrected by this method with a single third-strand binding site. It is preferred, however, that the third-strand binding site be within 10 kb  
30 of the genetic defect to be corrected.

In addition, the mutagen psoralen (for which mutagen-stimulated homologous recombination by the methods of this invention has been demonstrated and for which misrepair of T to A has been demonstrated) is used clinically for  
35 therapy --topically for dermatological conditions and ex vivo for bone marrow transplants to reduce the risk of graft-vs-host disease and as therapy for cutaneous T-cell

lymphoma. Also, psoralen alone is relatively non-toxic in clinical use (Ortonne, *Clin. Dermatol.* 7:120 (1989); Taylor and Gasparro, *Semin. Hematol.* 29:132 (1992); Jampel, et al., *Arch. Dermatol.* 127: 1673 (1991); Ullrich, *J. Invest. Dermatol.* 96:303 (1991)). In cell culture, psoralen-linked and acridine-linked third strands are less toxic than the drugs administered alone (Zerial, et al., *Nucleic Acids Res.* 15:9909 (1987)).

Further uses of the invention include restoration or destruction of  $\beta$  gene function in experimental cell lines and transgenic animals. A transgenic mouse expressing only human sickle-cell hemoglobin, for example, would be extremely useful for testing therapies for this disease in advance of human trials.

The use of and useful and novel features of mutagen-linked third strands to either cause a desired mutation (targeted mutagenesis) or cause DNA damage to stimulate homologous recombination with a donor DNA (third-strand-directed homologous recombination or targeted DNA replacement) will be further understood in view of the following non-limiting examples.

#### Example 1

A single mutation in the  $\beta$  chain of hemoglobin is sufficient to cause the sickle-shaped cells characteristic of sickle cell anemia (SCA). At the DNA level, the normal adenine base is mutated to a thymine, which causes an amino acid change from a negatively-charged glutamic acid to an uncharged, hydrophobic valine. The normal and altered DNA sequence surrounding and including the changed base (boldface) is:

Normal sequence:

gag

5' ATG GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG 3'  
3' GAC CAC GTG GAC TGA GGA CTC CTC TTC AGA CGG CAA TGA CGG GAC 5'  
(SEQ ID NO:1)

Altered sequence:

gug ②

5 5' ATG GTG CAC CTG ACT CCT GTG GAG AAG TCT GCC GTT ACT GCC CTG 3'  
 3' GAC CAC GTG GAC TGA GGA CAC CTC TTC AGA CGG CAA TGA CGG GAC 5'

① ③

(SEQ ID NO:2)

10 Or an alternative strand-switching possibility:

gug ⑤

5' ATG GTG CAC CTG ACT CCT GTG GAG AAG TCT GCC GTT ACT GCC CTG 3'  
 3' GAC CAC GTG GAC TGA GGA CAC CTC TTC AGA CGG CAA TGA CGG GAC 5'

④ ⑥

(SEQ ID NO:2)

In those sequences, the upper strand is the coding strand where coding begins at the first triplet shown, the ATG start codon for the  $\beta$  globin gene. The codons "gag" and "gug" code for the native glutamic acid and mutant valine, respectively. Relevant to the present invention are the facts that: (1) the altered DNA sequence is imbedded in nearly uninterrupted stretches of purine bases (underlined regions designated by the circled numbers ①, ②, ③ or ④, ⑤, ⑥) alternating between strands that are targets for binding a third strand with strand switching; and (2) in the targeted mutagenesis method (TM method) of this invention, highly-specific damage to thymine base by psoralen-bound third strands is preferentially "misrepaired" to adenine, just the change that is required to change HbS to normal HbA.

Utilizing the binding code and motifs for third-strand binding, example sequences of third strands that bind to this purine rich region are presented below. In all examples, oligonucleotides suitable for use in the present invention may be derived by any method known in the art,

including chemical synthesis, or by cleavage of a larger nucleic acid using non-specific nucleic acid-cleaving chemicals or enzymes, or by using site specific restriction endonucleases. Psoralen and other mutagens may also be specifically bound to the ends or internal positions of the oligonucleotides by standard methods (Havre, et al., *Proc. Natl. Acad. Sci. USA* 90:7879 (1993); Havre and Glazer, *J. Virology* 67:7324 (1993)).

Examples of sequences for the 5 motifs binding solely to the ② purine-rich region of the gene are:

1. Purine/antiparallel motif (beginning at base 19 from the 5' ATG initiation codon of the coding strand):

5' GAAGAGGNG 3' (SEQ ID NO:3; N=T, A, G or C)

The purine/antiparallel motif is the preferred purine motif for binding to the ② region alone (without strand switching) since it is a slightly G-rich target. The single pyrimidine base, T, in the ② region may be opposite an A, G, C or T base as indicated in the third strand depicted. While there are no strongly preferred bases for the mismatch, the T base is slightly preferred. That T base in the coding strand is also the one that it is desired to change to the native A base. The psoralen is preferably attached to the A, G, C or T base opposite the T base in the center strand.

Equally preferred is the slightly shorter third-strand sequence:

5' GAAGAGG 3' (SEQ ID NO:4)

That sequence binds to the ② region with an equilibrium dissociation constant (Kd) in the 10<sup>-5</sup> M range.

2. Purine/parallel motif (beginning at base 19 from the 5' end of the coding strand):

5' GNGGAGAAG 3' (SEQ ID NO:5; N=T, A, G or C)

The single pyrimidine base, T, in the ② region may be opposite an A, G, C or T base as indicated in the depicted third strand. That T base is also the one that it is desired to change to the native A base. The psoralen is preferably attached to the A, G, C or T base opposite the T base in the center strand.

10 3. Pyrimidine/parallel motif (beginning at base 19 from the 5' end of the coding sequence):

5' CNCCTCTTC 3' (SEQ ID NO:6; N=G, T, A or C)

15 The single pyrimidine base, T, in the ② region may be opposite an A, G, C or T base as indicated in the depicted third strand. For the pyrimidine/parallel motif the G base is the preferred base opposite the T base in the center strand in this sequence (see Table 3). That T base is also the one that it is desired to change to the native A base. The psoralen is preferably attached to the A, G, C or T base opposite the T base in the center strand.

25 4. T and G/antiparallel motif (beginning at base 19 from the 5' end of the coding sequence):

5' GTTGTGGNG 3' (SEQ ID NO:7; N=T, G, A or C)

30 The G and T/antiparallel motif is the slightly preferred of the two GT motifs for binding to ② region alone (without strand switching) since the target AG and GA nearest neighbor frequencies outnumber the AA and GG frequencies by 3 to 2. The single pyrimidine base, T, in the ② region may be opposite an A, G, C or T base as indicated in the depicted third strand. This T base is also the one that it is desired to change to the native A base. The psoralen is preferably attached to the A, G, C or T base opposite the T

base in the center strand.

5. T and G/parallel motif (beginning at base 19 from the 5' end of the coding strand):

5

5' GNGGTGTTG 3' (SEQ ID NO:8; N=T, A, G or C)

10

The single pyrimidine base, T, in the ② region may be opposite an A, G, C or T base as indicated in the depicted third strand. This T base is also the one that it is desired to change to the native A base. The psoralen is preferably attached to the A, G, C or T base opposite the T base in the center strand, or attached to the base immediately adjacent.

15

20

25

Examples of a number of preferred third-strand sequences utilizing strand switching and binding to all of the ①, ②, ③ purine-rich regions, and one example utilizing the ④, ⑤, ⑥ strand-switching scheme along with binding data, are presented below. It will be understood that fragments of these sequences utilizing the center purine-rich region and only one of the two adjacent regions are included in the description and within the scope of the invention although not explicitly shown. It is also understood that several combinations of motifs may be used to bind to the three regions although not explicitly shown.

30

1. The example immediately below employs the purine/antiparallel motif, throughout. It is the preferred purine motif, since the three center strand regions are G rich. In this example the ④, ⑤, ⑥ strand-switching scheme is utilized.

5'-GAGGATA3'-X-3'-GGAGAAG-5'-Y-5'-AGATGG-3'

35

X represents a linker (e.g., spacer phosphoroamidite 9 from Glen Research) required to both provide steric flexibility and to maintain antiparallel strand orientation after



strand switching. Y represents a linker required to both provide steric flexibility and to maintain antiparallel strand orientation after strand switching. In some binding experiments, the double-underlined sequence was omitted (see below). While there are no strongly preferred bases for the two mismatches, the T base is slightly preferred.

2. The example immediately below also employs the purine/antiparallel motif, throughout. It is the preferred purine motif, since the three center strand regions are G rich. In this and the following example, the ①, ②, ③ strand-switching scheme is utilized for illustrative purposes. All the following examples could also use the ④, ⑤, ⑥ scheme.

5'-GAGGA-3'-X-3'-GNGGAGAAG-5'-Y-5'-AGANGG-3' (N=T, C, G, A)

X represents a linker (e.g., spacer phosphoroamidite 9 from Glen Research) required to both provide steric flexibility and to maintain antiparallel strand orientation after strand switching. Y represents a linker required to both provide steric flexibility and to maintain antiparallel strand orientation after strand switching. While there are no strongly preferred bases for the two mismatches, the T base is slightly preferred.

3. The example immediately below employs the purine/antiparallel motif in regions ① and ③, and the pyrimidine/parallel motif in the ② region because in this motif a G-base mismatch opposite a T base in the center strand is a highly stable mismatch.

5'-GAGGA-Z-CNCCTCTTC-Z-GANGG-3' (first N=G, C, T, A; second N=T, C, G or A)

Since the third strand will have polarity 5' to 3'

throughout, the linker Z need only supply enough flexibility to make the switch. Such flexible linkers include, but are not limited to, one or two natural bases (e.g., T, TT, C, CC) and others such as spacer 3, spacer 9  
5 phosphoramidites from Glen Research (Sterling, VA).

4. The example immediately below employs the T and G/antiparallel motif in regions ① and ③ since it is slightly preferred for higher AG, GA nearest-neighbor  
10 frequencies, and the GT/parallel motif in the ② region, while slightly less preferred the third strand polarity remains 5' to 3', and with natural bases as flexible linkers (-TT- used in the example below) production at any scale will be simpler.

15

5' GTGGT-TT-CNCCTCTTC-TT-GTNGG 3'

(SEQ ID NO:9; first N=G, C, T, A; second N=T, C, G or A).

It is understood that there are a number of acceptable third strand sequence, motif, and strand polarity  
20 combinations, which are within the scope of this invention although not explicitly listed.

#### Example 2

The  $\beta$ -gene cluster, shown in Figure 3, is located on  
25 human chromosome 11 and contains all the  $\beta$ -like genes in the order they are expressed in human development, from left to right in the Figure. The cluster from the beginning of the  $\epsilon$  gene to the end of the  $\beta$  gene spans about 52 kilobases (Kb). The targeted DNA replacement  
30 method (TDR), allows for DNA modification or damage to occur at greater than 52 Kb from the site at which a desired DNA change is to be made, so one third strand binding site may be used to repair any genetic defect or make any other alteration in the whole  $\beta$ -gene cluster. It  
35 is preferred, however, that the DNA damage site be within 10 kb of the site of the repair or alteration. For some clinical applications, it may be preferable that the DNA-

damage site be even closer to the site at which the DNA is to be altered.

In the  $\beta$  globin gene sequence, particularly in the introns, there are many good third-strand binding sites that may be utilized in the practice of the TDR method of the invention. A portion of the GenBank sequence of the chromosome-11 human-native hemoglobin-gene cluster (GenBank: LOCUS HUMHBB, 73308 bp ds-DNA) from base 60001 to base 66060 is presented below. This portion of the GenBank sequence contains the native  $\beta$  globin gene sequence. In sickle cell hemoglobin the adenine base at position 62206 (or position 2206 as listed in SEQ ID NO:10, indicated in boldface) is mutated to a thymine. The start of the gene coding sequence at position 62187-62189 (or positions 2187-2189 of SEQ ID NO:10) is indicated by a single underline.

A computer search was performed on this GenBank sequence portion for third-strand binding sites, and a representative sample of sites found are indicated by double-underlines. The preferred sites for the TDR method of this invention are both double-underlined and boldface.

	AAAGCTCTTG	CTTTGACAAT	TTTGGTCTTT	CAGAATACTA	TAAATATAAC	50
	CTATATTATA	ATTTCATAAA	GTCTGTGCAT	TTTCTTTGAC	CCAGGATATT	100
	TGCAAAAGAC	ATATTCAAAC	TTCCGCAGAA	CACTTTATTT	CACATATACA	150
25	TGCCTCTTAT	ATCAGGGATG	TGAAACAGGG	TCTTGAAAAC	TGTCTAAATC	200
	TAAAACAATG	CTAATGCAGG	TTTAAATTTA	ATAAAATAAA	ATCCAAAATC	250
	TAACAGCCAA	GTCAAATCTG	TATGTTTTTA	CATTTAAAT	ATTTTAAAGA	300
	CGTCTTTTCC	CAGGATTCAA	CATGTGAAAT	CTTTTCTCAG	GGATACACGT	350
	GTGCCTAGAT	CCTCATTGCT	TTAGTTTTTT	ACAGAGGAAT	GAATATAAAA	400
30	AGAAAATACT	TAAATTTTAT	CCCTCTTACC	TCTATAATCA	TACATAGGCA	450
	TAATTTTTTA	ACCTAGGCTC	CAGATAGCCA	TAGAAGAACC	AAACACTTTC	500
	TGCGTGTGTG	AGAATAATCA	GAGTGAGATT	TTTTCACAAG	TACCTGATGA	550
	GGGTTGAGAC	AGGTAGAAAA	AGTGAGAGAT	CTCTATTTAT	TTAGCAATAA	600
	TAGAGAAAGC	ATTTAAGAGA	ATAAAGCAAT	GGAAATAAGA	AATTTGTAAA	650
35	TTTCCTTCTG	ATAACTAGAA	ATAGAGGATC	CAGTTTCTTT	TGGTTAACCT	700
	AAATTTTATT	TCATTTTATT	GTTTTATTTT	ATTTTATTTT	ATTTTATTTT	750
	GTGTAATCGT	AGTTTCAGAG	TGTTAGAGCT	<u>GAAAGGAAGA</u>	<u>AGTAGGAGAA</u>	800

ACATGCAAAG TAAAAGTATA ACACTTTCCT TACTAAACCG ACTGGGTTTC 850  
 CAGGTAGGGG CAGGATTCAG GATGACTGAC AGGGCCCTTA GGGAACACTG 900  
 AGACCCTACG CTGACCTCAT AAATGCTTGC TACCTTTGCT GTTTTAATTA 950  
 CATCTTTTAA TAGCAGGAAG CAGA ACTCTG CACTTCAAAA GTTTTTCCTC 1000  
 5 ACCTGAGGAG TTAATTTAGT ACAAGGGGAA AAAGTACAGG GGGATGGGAG 1050  
 AAAGGCGATC ACGTTGGGAA GCTATAGAGA AAGAAGAGTA AATTTTAGTA 1100  
 AAGGAGGTTT AAACAAACAA AATATAAAGA GAAATAGGAA CTTGAATCAA 1150  
 GGAAATGATT TTAACACGCA GTATTCTTAG TGGACTAGAG GAAAAAATA 1200  
 ATCTGAGCCA AGTAGAAGAC CTTTTCCCCT CCTACCCCTA CTTTCTAAGT 1250  
 10 CACAGAGGCT TTTTGTTCCT CCAGACACTC TTGCAGATTA GTCCAGGCAG 1300  
 AAACAGTTAG ATGTCCCCAG TTAACCTCCT ATTTGACACC ACTGATTACC 1350  
 CCATTGATAG TCACACTTTG GGTGTGTAAGT GACTTTTTTAT TTATTTGTAT 1400  
 TTTTGACTGC ATTAAGAGGT CTCTAGTTTT TTATCTCTTG TTTCCCAAAA 1450  
 CCTAATAAGT AACTAATGCA CAGAGCACAT TGATTTGTAT TTATCTATT 1500  
 15 TTTAGACATA ATTTATTAGC ATGCATGAGC AAATTAAGAA AAACAACAAC 1550  
 AAATGAATGC ATATATATGT ATATGTATGT GTGTATATAT ACACATATAT 1600  
 ATATATATTT TTTTTCTTTT CTTACCAGAA GGTTTTAATC CAAATAAGGA 1650  
 GAAGATATGC TTAGAACTGA GGTAGAGTTT TCATCCATTC TGTCTGTAA 1700  
 GTATTTTGCA TATTCTGGAG ACGCAGGAAG AGATCCATCT ACATATCCCA 1750  
 20 AAGCTGAATT ATGGTAGACA AAGCTCTTCC ACTTTTAGTG CATCAATTTT 1800  
 TTATTTGTGT AATAAGAAAA TTGGGAAAC GATCTTCAAT ATGCTTACCA 1850  
 AGCTGTGATT CCAAATATTA CGTAAATACA CTTGCAAAGG AGGATGTTTT 1900  
 TAGTAGCAAT TTGTACTGAT GGTATGGGGC CAAGAGATAT ATCTTAGAGG 1950  
 GAGGGCTGAG GGTTTGAAGT CCAACTCCTA AGCCAGTGCC AGAAGAGCCA 2000  
 25 AGGACAGGTA CGGCTGTCAT CACTTAGACC TCACCCTGTG GAGCCACACC 2050  
 CTAGGGTTGG CCAATCTACT CCCAGGAGCA GGGAGGGCAG GAGCCAGGGC 2100  
 TGGGCATAAA AGTCAGGGCA GAGCCATCTA TTGCTTACAT TTGCTTCTGA 2150  
 CACAACTGTG TTTACTAGCA ACCTCAAACA GACACCATGG TGCACCTGAC 2200  
 TCCTGAGGAG AAGTCTGCCG TTTACTGCCCT GTGGGGCAAG GTGAACGTGG 2250  
 30 ATGAAGTTGG TGGTGAGGCC CTGGGCAGGT TGGTATCAAG GTTACAAGAC 2300  
 AGGTTTAAGG AGACCAATAG AAAGTGGGCA TGTGGAGACA GAGAAGACTC 2350  
 TTGGGTTTCT GATAGGCACT GACTCTCTCT GCCTATTGGT CTATTTTCCC 2400  
 ACCCTTAGGC TGCTGGTGGT CTACCCTTGG ACCCAGAGGT TCTTTGAGTC 2450  
 CTTTGGGGAT CTGTCCACTC CTGATGCTGT TATGGGCAAC CCTAAGGTGA 2500  
 35 AGGCTCATGG CAAGAAAGTG CTCGGTGCCT TTAGTGATGG CCTGGCTCAC 2550  
 CTGGACAACC TCAAGGGCAC CTTTGCCACA CTGAGTGAGC TGCACCTGTGA 2600  
 CAAGCTGCAC GTGGATCCTG AGAACTTCAG GGTGAGTCTA TGGGACCCTT 2650

GATGTTTTCT TTCCCCTTCT TTTCTATGGT TAAGTTCATG TCATAGGAAG 2700  
 GGGAGAAGTA ACAGGGTACA GTTTAGAAATG GGAAACAGAC GAATGATTGC 2750  
 ATCAGTGTGG AAGTCTCAGG ATCGTTTTAG TTTCTTTTAT TTGCTGTTCA 2800  
 TAACAATTGT TTTCTTTTGT TTAATTCTTG CTTCTTTTT TTTTCTTCTC 2850  
 5 CGCAATTTTT ACTATTATAC TTAATGCCTT AACATTGTGT ATAACAAAAG 2900  
 GAAATATCTC TGAGATACAT TAAGTAACTT AAAAAAAAC TTTACACAGT 2950  
 CTGCCTAGTA CATTACTATT TGGAATATAT GTGTGCTTAT TTGCATATTC 3000  
 ATAATCTCCC TACTTTATTT TCTTTTATTT TTAATTGATA CATAATCATT 3050  
 ATACATATTT ATGGGTTAAA GTGTAATGTT TTAATATGTG TACACATATT 3100  
 10 GACCAAATCA GGGTAATTTT GCATTTGTAA TTTTAAAAA TGCTTTCTTC 3150  
 TTTTAATATA CTTTTTGTGTT TATCTTATTT CTAATACTTT CCCTAATCTC 3200  
TTTCTTTCAG GGCAATAATG ATACAATGTA TCATGCCTCT TTGCACCATT 3250  
 CTAAAGAATA ACAGTGATAA TTTCTGGGTT AAGGCAATAG CAATATTTCT 3300  
 GCATATAAAT ATTTCTGCAT ATAAATTGTA ACTGATGTAA GAGGTTTCAT 3350  
 15 ATTGCTAATA GCAGCTACAA TCCAGCTACC ATTCTGCTTT TATTTTATGG 3400  
 TTGGGATAAG GCTGGATTAT TCTGAGTCCA AGCTAGGCCC TTTTGCTAAT 3450  
 CATGTTTATA CCTCTTATCT TCCTCCACAA GCTCCTGGGC AACGTGCTGG 3500  
 TCTGTGTGCT GGCCCATCAC TTTGGCAAAG AATTCACCCC ACCAGTGCAG 3550  
 GCTGCCTATC AGAAAGTGGT GGCTGGTGTG GCTAATGCCC TGGCCACAAA 3600  
 20 GTATCACTAA GCTCGCTTTC TTGCTGTCCA ATTTCTATTA AAGGTTCCCTT 3650  
 TGTTCCCTAA GTCCAACCTAC TAAACTGGGG GATATTATGA AGGGCCTTGA 3700  
 GCATCTGGAT TCTGCCTAAT AAAAAACATT TATTTTCATT GCAATGATGT 3750  
 ATTTAAATTA TTTCTGAATA TTTTACTAAA AAGGGAATGT GGGAGGTCAG 3800  
 TGCATTTAAA ACATAAAGAA ATGAAGAGCT AGTTCAAACC TTGGGAAAAT 3850  
 25 ACACTATATC TTAAACTCCA TGAAAGAAGG TGAGGCTGCA AACAGCTAAT 3900  
 GCACATTGGC AACAGCCCTG ATGCCTATGC CTTATTCATC CCTCAGAAAA 3950  
 GGATTCAAGT AGAGGCTTGA TTTGGAGGTT AAAGTTTTGC TATGCTGTAT 4000  
 TTTACATTAC TTATTGTTTT AGCTGTCTCT ATGAATGTCT TTTCACTACC 4050  
 CATTTGCTTA TCCTGCATCT CTCAGCCTTG ACTCCACTCA GTTCTCTTGC 4100  
 30 TTAGAGATAC CACCTTTCCC CTGAAGTGTT CCTTCCATGT TTTACGGCGA 4150  
 GATGGTTTCT CCTCGCCTGG CCACTCAGCC TTAGTTGTCT CTGTTGTCTT 4200  
 ATAGAGGTCT ACTTGAAGAA GGAAAAACAG GGGGCATGGT TTGACTGTCC 4250  
 TGTGAGCCCT TCTTCCCTGC CTCCCCCACT CACAGTGACC CGGAATCTGC 4300  
 AGTGCTAGTC TCCCGGAAC ATCACTCTTT CACAGTCTGC TTTGGAAGGA 4350  
 35 CTGGGCTTAG TATGAAAAGT TAGGACTGAG AAGAATTGA AAGGGGGCTT 4400  
 TTTGTAGCTT GATATTCAT ACTGTCTTAT TACCCTATCA TAGGCCACC 4450

CCAAATGGAA GTCCCATTTCT TCCTCAGGAT GTTTAAGATT AGCATTCAGG 4500  
 AAGAGATCAG AGGTCTGCTG GCTCCCTTAT CATGTCCCTT ATGGTGCTTC 4550  
 TGGCTCTGCA GTTATTAGCA TAGTGTTACC ATCAACCACC TTAAC TTCAT 4600  
 TTTTCTTATT CAATACCTAG GTAGGTAGAT GCTAGATTCT GGAAATAAAA 4650  
 5 TATGAGTCTC AAGTGGTCCT TGTCTCTCTC CCCAGTCAAA TTCTGAATCT 4700  
 AGTTGGCAAG ATTCTGAAAT CAAGGCATAT AATCAGTAAT AAGTGATGAT 4750  
 AGAAGGGTAT ATAGAAGAAT TTTATTATAT GAGAGGGTGA AACCTAAAAT 4800  
 GAAATGAAAT CAGACCCTTG TCTTACACCA TAAACAAAAA TAAATTTGAA 4850  
 TGGGTAAAG AATTAAACTA AGACCTAAAA CCATAAAAAT TTTTAAAGAA 4900  
 10 ATCAAAAGAA GAAAATTCTA ATATTCATGT TGCAGCCGTT TTTTGAATTT 4950  
 GATATGAGAA GCAAAGGCAA CAAAAGGAAA AATAAAGAAG TGAGGCTACA 5000  
 TCAAACTAAA AAATTTCCAC ACAAAAAGA AAACAATGAA CAAATGAAAG 5050  
 GTGAACCATG AAATGGCATA TTTGCAAACC AAATATTTCT TAAATATTTT 5100  
 GGTAAATATC CAAAATATAT AAGAAACACA GATGATTCAA TAACAAACAA 5150  
 15 AAAATTAAAA ATAGGAAAAT AAAAAATTA AAAAGAAGAA AATCCTGCCA 5200  
 TTTATGCGAG AATTGATGAA CCTGGAGGAT GTAAAACTAA GAAAAATAAG 5250  
 CCTGACACAA AAAGACAAAT ACTACACAAC CTTGCTCATA TGTGAAACAT 5300  
 AAAAAAGTCA CTCTCATGGA AACAGACAGT AGAGGTATGG TTTCCAGGGG 5350  
 TTGGGGGTGG GAGAATCAGG AACTATTAC TCAAAGGGTA TAAAATTTCA 5400  
 20 GTTATGTGGG ATGAATAAAT TCTAGATATC TAATGTACAG CATCGTGA CT 5450  
 GTAGTTAATT GTACTGTAAG TATATTTAAA ATTTGCAAAG AGAGTAGATT 5500  
 TTTTTGTTTT TTTAGATGGA GTTTTGCTCT TGTTGTCCAG GCTGGAGTGC 5550  
 AATGGCAAGA TCTTGGCTCA CTGCAACCTC CGCCTCCTGG GTTCAAGCAA 5600  
 ATCTCCTGCC TCAGCCTCCC GAGTAGCTGG GATTACAGGC ATGCGACACC 5650  
 25 ATGCCCAGCT AATTTTGAT TTTTAGTAGA GACGGGGTTT CTCCATGTTG 5700  
 GTCAGGCTGA TCCGCCTCCT CGGCCACCAA AGGGCTGGGA TTACAGGCGT 5750  
 GACCACCGGG CCTGGCCGAG AGTAGATCTT AAAAGCATTT ACCACAAGAA 5800  
 AAAGGTA ACT ATGTGAGATA ATGGGTATGT TAATTAGCTT GATTGTGGTA 5850  
 ATCATTTTCAC AAGGTATACA TATATTAAAA CATCATGTTG TACACCTTAA 5900  
 30 ATATATACAA TTTTATTTG TGAATGATAC CTCAATAAAG TTGAAGAATA 5950  
 ATAAAAAGA ATAGACATCA CATGAATTAA AAAACTAAAA AATAAAAAAA 6000  
 TGCATCTTGA TGATTAGAAT TGCATTCTTG ATTTTTCAGA TACAAATATC 6050  
 CATTTGACTG 6060  
 (SEQ ID NO:10)

35

It is understood that these third-strand binding sites

are illustrative and do not constitute all the sites in the region, which are also within the scope of the invention.

The two preferred binding sites, beginning at Gen Bank positions 62655 and 62825 (SEQ ID NO:10 positions 2655 and 2825), are each 21 uninterrupted pyrimidines in the coding strand or 21 uninterrupted purines in the non-coding strand and are excellent third-strand binding sites. Their sequences are:

10 TTTTCTTTCC CCTTCTTTTC T (SEQ ID NO:11) and  
CTTTCTTTTT TTTTCTTCTC C (SEQ ID NO:12)

Both are located in the second intron or intervening sequence (IVS-2) of the  $\beta$  globin gene.

15 To illustrate third-strand compositions of matter that will bind tightly to those sites in the practice of the TDR method of the invention, we choose the first of the two sites. The sequence in double-stranded form is:

20 5' TTTTCTTTCC CCTTCTTTTC TA 3' (coding strand) (SEQ ID NO:13)  
3' AAAAGAAAGG GGAAGAAAAG AT 5'

Since the purine run on the non-coding strand is 21 bases long and is not interrupted by even one pyrimidine, it exceeds the preferred minimum length of 20 bases for third strand binding. The A:T base pair at the 3' end of the coding strand after the purine run are shown because it represents a good crosslinking site for psoralen attached to the end of the third strand. The site is also conveniently located to cause DNA damage to stimulate homologous recombination using a donor DNA carrying desired alterations to coding regions and introns of the  $\beta$ -gene and to adjacent control regions of the  $\beta$ -gene. While in the invention DNA base positions to be altered by the donor DNA are not required to be so near the site of DNA damage, a third-strand binding sequence located in the  $\beta$  globin gene

allows for flexibility in the length of donor DNA, which may be used to optimize introduction into hematopoietic stem cells or erythrocyte precursor cells, optimize homologous recombination, or allow for donor DNA of  
5 sufficiently short length to be delivered by traditional means, injection or IV, to a patient.

According to the binding code and motifs for third-strand binding, example sequences of third strands that bind to this purine rich region are presented below. In  
10 all examples, oligonucleotides suitable for use in the present invention may be derived by any method known in the art, including chemical synthesis, or by cleavage of a larger nucleic acid using non-specific nucleic acid-cleaving chemicals or enzymes, or by using site specific  
15 restriction endonucleases. Psoralen and other mutagens may also be specifically bound to the ends or internal positions of the oligonucleotides by standard methods. Donor DNA may be prepared in the same manner. The example sequences are 21 bases and consequently bind to the entire  
20 purine run. It is understood that effective fragments thereof are included within the scope of the invention.

1. The purine motif example immediately below, employs the parallel polarity, which is preferred because  
25 the target is A rich.

5' psoralen-AGAAAAGAAG GGGAAAGAAA A 3' (SEQ ID NO:14)

While not illustrated, the antiparallel motif may be  
30 employed, although not preferred. Psoralen crosslinking to the AT base pair at the end of the target is a preferred method of causing DNA damage, and the position where it is bound is illustrated above. It is understood that other mutagens and other positions for binding to third strands  
35 are within the scope of the invention. The examples below, therefore, will not illustrate mutagen binding.



2. The T and G motif example immediately below, employs the parallel polarity, which is preferred because the target has high AA and GG nearest neighbor frequencies.

5 5' TGTTTGTG GGGTTGTTT T 3' (SEQ ID NO:15)

While not illustrated, the antiparallel motif may be employed, although not preferred.

10 3. A third strand in the pyrimidine/parallel motif is another example within the scope of the invention:

5' TCTTTTCTTC CCCTTTCTTT T 3' (SEQ ID NO:16)

15 4. Mixed motifs are also within the scope of the invention. A mixed purine/parallel and pyrimidine/parallel motif third strand which will bind to the target in question is illustrated immediately below:

20 5' TCTTTTCTTG GGGAAAGAAA A 3' (SEQ ID NO:17)

The donor DNA must contain the DNA sequence at the DNA damage site, the DNA region containing the genetic defect to be repaired or alteration to be made, and all the native codons between the two (preferably 50 or more bases of homology to the target DNA). The donor DNA may be considerably larger than the bases between and including the damage site and the repair or alteration site. For repair of sickle cell anemia to native DNA and protein sequence, for example, the donor DNA must contain both the native adenine that is to replace the mutant thymine, the third-strand binding site to be damaged, and preferably the native DNA sequence between them. One example of a double-stranded donor DNA meeting these requirements is presented immediately below (only one strand of the duplex DNA is shown). The strand spans positions 62161-62760 of GenBank: LOCUS HUMHBB, or positions 2161-2760 of (SEQ ID NO:10):

TTCACTAGCA ACCTCAAACA GACACCATGG TGCACCTGAC TCCTGAGGAG 50  
 AAGTCTGCCG TTACTGCCCT GTGGGGCAAG GTGAACGTGG ATGAAGTTGG 100  
 TGGTGAGGCC CTGGGCAGGT TGGTATCAAG GTTACAAGAC AGGTTTAAGG 150  
 5 AGACCAATAG AAACCTGGGCA TGTGGAGACA GAGAAGACTC TTGGGTTTCT 200  
 GATAGGCACT GACTCTCTCT GCCTATTGGT CTATTTTCCC ACCCTTAGGC 250  
 TGCTGGTGGT CTACCCTTGG ACCCAGAGGT TCTTTGAGTC CTTTGGGGAT 300  
 CTGTCCACTC CTGATGCTGT TATGGGCAAC CCTAAGGTGA AGGCTCATGG 350  
 CAAGAAAGTG CTCGGTGCCT TTAGTGATGG CCTGGCTCAC CTGGACAACC 400  
 10 TCAAGGGCAC CTTTGCCACA CTGAGTGAGC TGCACGTGTA CAAGCTGCAC 450  
 GTGGATCCTG AGAACTTCAG GGTGAGTCTA TGGGACCCTT GATGTTTTCT 500  
TTCCCCTTCT TTTCTATGGT TAAGTTCATG TCATAGGAAG GGGAGAAGTA 550  
 ACAGGGTACA GTTTAGAATG GGAAACAGAC GAATGATTGC ATCAGTGTGG 600  
 (SEQ ID NO:18)

15

It is understood by those skilled in the art that this is only one example of a very large number of suitable donor nucleic acids. Some variations would include but are not limited to: longer and shorter donor DNAs that contain both the native A base at position 62206 and the third-strand binding site; donor DNAs with different codons that code for the same amino acid or a mutant amino acid that codes for a functional protein; donor DNAs with sequence variations in the introns that do not effect substantially the processing of protein.

25

### Example 3

Mutations causing  $\beta^0$  and  $\beta^+$  thalassemia are mostly found in the  $\beta$  globin gene itself or in DNA regions close to the gene. A complete list of mutations, as of 1993, may be found in Huisman (Hemoglobin, 17:479 (1993)). Some examples of mutations causing  $\beta^0$  thalassemia from Huisman and for which the donor DNA of Example 2 may be used in the TDR method to repair the mutation are presented below. It is understood that all the mutations listed in Huisman and those yet to be discovered that are located at sequence positions within the region spanned by the donor DNA may be

35

repaired using the donor DNA.

1. RNA processing mutants at splice junctions:

5 G-->A at IVS-1-1 (that is, the G (normal base) to A  
(mutation) in intervening sequence 1 at position 1 in the  
intervening sequence). Using this same shorthand notation,  
others are:

10 G-->T at IVS-1-1  
T-->C at IVS-1-2  
17 nucleotide deletion at 3' end of IVS-1

2. Nonsense (stop) mutants in coding regions

15 TGG --> TAG at codon 15  
AAG-->TAG at codon 17

3. Frameshift mutants in the coding region

20 CCT-->C-- at codon 5  
GTT-->GT- at codon 11  
C inserted between codons 9 and 10

25 4. Initiation codon mutations

ATG-->ACG at codon 1  
ATG-->AGG at codon 1  
ATG-->GTG at codon 1

30

5. Hyperunstable or non-functional hemoglobins

CTG-->CGG at codon 28 (Leu-->Arg)  
CTG GTG --> CT- --G deletion in codons 32 and 33

35

Example 4

Husiman (Hemoglobin, 17:479 (1993)) lists greater than 60 mutations for  $\beta^0$  thalassemia alone, and lists greater than 45 mutations for  $\beta^+$  thalassemia. All these defective  $\beta$  chains may be corrected in the TDR method using donor DNAs which are comprised of the DNA region spanning the genetic defect and the third-strand binding site in IVS-2 of Example 2. Such donor DNAs are simply the native DNA sequence spanning the region between the third-strand binding site and the site of the mutation to be repaired, preferably 50 or more bases.

Example 5

The TM method may also be used to correct  $\beta$  thalassemia provided: (1) the mutation lies in a third strand binding site, and (2) repair of the mutation yields either a native or non-native base that results in a functional hemoglobin. Frequent base changes from the action of specific mutagens may be found in: Aguilar, et al., *Proc. Natl. Acad. Sci. USA.* 90: 8586 (1993); Gupta, et al., *J. Biol. Chem.* 264: 20120 (1989); Topal, *Carcinogenesis*. 9: 691 (1988); Moriya, *Proc. Natl. Acad. Sci. USA.* 90: 1122 (1993); Klein, et al., *Nucleic Acids Res.* 18: 4131 (1990).

Of particular interest are nonsense mutations (and mutations in the initiation codon ATG), since a single base change produces a shortened hemoglobin protein (or no hemoglobin) that will almost certainly be nonfunctional resulting in the severe  $\beta^0$  thalassemia condition. From a table of the genetic code, the three stop codons are TAA, TAG, and TGA. In the table below are example mutagens, the base changes that they cause most frequently, and the amino acid resulting from that base change in a stop codon.

Stop Codon	Mutagen (frequent base change)	Resulting Codons	Resulting Amino Acids
TAA	psoralen (T-->A)	AAA	Lys
TAG	psoralen (T-->A)	AAG	Lys
	alkylating agents (A-->G, G-->T)	TGG, TAT	Trp, Tyr
	MNU [N-methyl- N nitrosourea] (G-->C)	TAC	Tyr
TGA	psoralen (T-->A)	AGA	Arg
	alkylating agents (A-->G, G-->T)	TGG, TTA	Trp, Leu
	MNU (G-->C)	TCA	Ser

Some mutagens can change more than one base or cause more than one type of mutation, depending on the location of the mutagen on the third strand (i.e., what base in the duplex the mutagen is near upon third strand binding), nearest neighbor bases in the duplex, and other factors.

For  $\beta^0$  thalassemia caused by a CAG--->TAG mutation in the codon  $\beta$ 39 that results in a Gln---> Stop (nonsense) codon change, a psoralen-linked third strand targeted to the T base can change this nonsense codon to AAG which codes for lysine to provide a functional  $\beta$  chain and hemoglobin (S. Baserga, Ph.D. Thesis, Yale University (1988)). This site also fulfills the second requirement of the TM method, as it is a weak third-strand binding site using strand switching. The sequence at the site spanning codons 39 through 42 in the native,  $\beta^0$  thalassemia, and

psoralen-linked third-strand repaired is:

39 40 41 42

5 5' CAG AGG TTC TTT 3' (normal) (SEQ ID NO:19)  
 5' TAG AGG TTC TTT 3' ( $\beta^0$  thalassemia) (SEQ ID NO:20)  
 5' AAG AGG TTC TTT 3' (repaired) (SEQ ID NO:21)

10 The first underlined sequence in the  $\beta^0$  thalassemia DNA is the first purine run for third-strand binding, and the second underlined sequence is the complement to the AAGAAA sequence to which the third-strand will bind after switching.

15 Three preferred third-strands that will bind to that portion of the  $\beta^0$  thalassemia sequence, where the psoralen is represented as **pso**, are:

20 **pso**  
 |  
 3'-AGAGG-Z-AAGAAA-5'

25 **pso**  
 |  
 3'-TGTGG-Z-TTGTTT-5'

30 **pso**  
 |  
 3'-AGAGG-Z-TTCTTT-5'

Longer antiparallel third strands may be employed that utilizes additional strand switches. One example is:

35 **pso**  
 |  
 3'-AGAGG-Z-TTCTTT-Z-GAG-Z-TCCTTT-Z-GGGGA-5'

In these example third strands, Z is a linker to provide flexibility for strand switching but does not change the polarity of the third strand.

5 Other thalassemias are caused by improper processing of introns (intervening sequences or IVS). Of particular interest, at the 3' ends of introns the consensus sequence is

10 (T or C)<sub>n</sub>N(C or T)AGG-3'

where n>10, N stands for any nucleotide, and the underlined AG sequence is invariant among many species, and therefore, is thought to be absolutely required for proper splicing (Bunn and Forget, *op. cit.*, pp. 177-178). Relevant to this invention, the presence of the (T or C)<sub>n</sub> sequence provides a third-strand polypurine target site on the opposite strand for repairing improper splicing by the TM method.

15 In the  $\beta$  gene, the sequences at the ends of IVS-1 and IVS-2 (before the beginnings of exon 2 and exon 3) are:

5' CTCTCTCTGC CTATTGGTCT ATTTTCCCAC CCTTAGC exon 2  
(SEQ ID NO:22)

25 5'...CCTCTTATCT TCCTCCCACAGC exon 3 (SEQ ID NO:23)

The opposite strands of both these sequences represent excellent third-strand binding sites with few mismatches and without strand switching, but in some cases switching the polarity of the third strand according to the preferred motifs is desirable. The double-underlined AG bases represent the usually invariant bases necessary for proper splicing of the mRNA. For example, at the beginning of IVS-1, a G to A change at position 1 causes aberrant splicing in a Mediterranean Thalassemia (Orkin, *et al.*, *Nature* 296:627, 1982).

30  
35

These examples illustrate the concept of converting a stop (nonsense) codon to a non-native amino acid using mutagen-linked third strands in the TM method, which yields a functional  $\beta$  chain.

5



## WHAT IS CLAIMED IS:

1. A method for effecting homologous recombination between a native nucleic acid segment in a cell and a donor nucleic acid segment introduced into the cell, which  
5 comprises:

a) introducing into a human cell i) an oligonucleotide third strand which comprises a base sequence capable of forming a triple helix at a binding  
10 region on one or both strands of a native nucleic acid segment, said native nucleic acid segment containing an undesired mutation in the vicinity of the human  $\beta$  globin gene cluster target region where the recombination is to occur, said oligonucleotide being capable of inducing  
15 homologous recombination at the target region of the native nucleic acid, and ii) a donor nucleic acid which comprises a nucleic acid sequence sufficiently homologous to the native nucleic acid segment such that the donor sequence is capable of undergoing homologous recombination with the  
20 native sequence at the target region;

b) allowing the oligonucleotide third strand to bind to the native nucleic acid segment to form a triple stranded nucleic acid, thereby inducing homologous recombination at the native nucleic acid segment target  
25 region; and

c) allowing homologous recombination to occur between the native and donor nucleic acid segments.

2. The method of claim 1, wherein the oligonucleotide  
30 third strand is from about 7 to about 50 nucleotides in length.

3. The method of claim 2, wherein the oligonucleotide third strand is from about 10 to about 30 nucleotides in  
35 length.

4. The method of claim 1, wherein the oligonucleotide

third strand contains an at least partially artificial backbone.

5        5. The method of claim 1, wherein the oligonucleotide third strand contains a backbone selected from the group consisting of phosphodiester, phosphorothioate, methyl phosphonate, peptide, and mixtures thereof.

10       6. The method of claim 5, wherein the backbone is phosphodiester.

15       7. The method of claim 1, wherein the oligonucleotide third strand is modified with one or more protective groups.

      8. The method of claim 7, wherein the 3' and 5' ends of the oligonucleotide third strand are modified with one or more protective groups.

20       9. The method of claim 7, wherein the protective group is selected from the group consisting of alkyl amines, thiols, cholesterol, acridine and psoralen.

25       10. The method of claim 1, wherein the oligonucleotide third strand is circularized.

30       11. The method of claim 1, wherein the oligonucleotide third strand contains at least one modified sugar.

      12. The method of claim 1, wherein the oligonucleotide third strand has linked thereto a moiety which induces the homologous recombination.

35       13. The method of claim 12, wherein the moiety is linked to the oligonucleotide directly.

14. The method of claim 12, wherein the moiety is linked to the oligonucleotide through a linker.

5 15. The method of claim 12, wherein the moiety is selected from the group consisting of psoralen, a substituted psoralen, hydroxymethylpsoralen, mitomycin C, 1-nitrosopyrene, a nuclease, a restriction enzyme, a radionuclide, boron, and iodine.

10 16. The method of claim 1, wherein the donor nucleic acid is double stranded.

15 17. The method of claim 1, wherein the donor nucleic acid is single stranded.

18. The method of claim 1, wherein the donor nucleic acid comprises two substantially complementary single strands.

20 19. The method of claim 1, wherein the donor nucleic acid is substantially homologous with the native nucleic acid.

25 20. The method of claim 19, wherein the donor nucleic acid is substantially homologous with the native nucleic acid in a region of about 20 bases at each end of the donor nucleic acid.

30 21. The method of claim 1, wherein the donor nucleic acid is at least about 40 bases in length.

22. The method of claim 21, wherein the donor nucleic acid is between about 40 and about 40,000 bases in length.

35 23. The method of claim 1, wherein the donor nucleic acid is introduced into the cell in the form of a packaging system.

24. The method of claim 23, wherein the packaging system is selected from the group consisting of a DNA virus, an RNA virus, and a liposome.

5

25. The method of claim 1, wherein the cells are treated *in vivo*.

26. The method of claim 1, wherein the cells are hematopoietic stem cells which are treated *ex vivo*.

10

27. The method of claim 1, wherein the homologous recombination causes an alteration in the native nucleic acid sequence.

15

28. The method of claim 27, wherein the alteration is an addition of a segment selected from the group consisting of a gene, a part of a gene, a gene control region, an intron, a splice junction, a transposable element, a site specific recombination sequence, and combinations thereof.

20

29. The method of claim 1, wherein the oligonucleotide third strand comprises at least one synthetic base.

25

30. The method of claim 1, wherein the oligonucleotide third strand has a dissociation constant for the binding region of less than or equal to about  $10^{-7}$  M.

30

31. The method of claim 30, wherein the dissociation constant is less than or equal to about  $2 \times 10^{-8}$ .

32. The method of claim 1, wherein the native nucleic acid binding region comprises a sequence of at least seven bases contained within SEQ ID NO:10.

35

33. The method of claim 1, wherein the native nucleic acid target region comprises a sequence of at least seven bases contained within SEQ ID NO:10.

5           34. The method of claim 32, wherein the native nucleic acid binding region comprises a sequence of at least seven bases contained in one of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:23.

10           35. The method of claim 1, wherein the oligonucleotide is selected from the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17.

15           36. A method for effecting homologous recombination between a first nucleic acid segment in a cell and a donor nucleic acid segment introduced into the cell, which comprises:

20           a) contacting a donor nucleic acid segment with an oligonucleotide third strand which comprises a base sequence capable of forming a triple helix at a binding region on one or both strands of the donor nucleic acid segment in the vicinity of a target region where the  
25           recombination is to occur, said oligonucleotide being capable of inducing homologous recombination at the target region of the donor nucleic acid, and said donor having a sequence sufficiently homologous to a first nucleic acid segment in a human cell which comprises at least a portion  
30           of the human  $\beta$  globin gene cluster, such that the donor sequence will undergo homologous recombination with the first sequence at the target region;

35           b) treating the nucleic acid segment by allowing the oligonucleotide to bind to the donor nucleic acid segment to form a triple stranded nucleic acid;

          c) introducing into a human cell the treated donor nucleic acid; and

d) allowing homologous recombination to occur between the first and donor nucleic acid segments.

37. A composition comprising:

5 a) an oligonucleotide third strand which comprises a base sequence which is capable of forming a triple helix at a binding region on one or both strands of a native nucleic acid segment in the vicinity of the human  $\beta$  globin gene cluster target region, said oligonucleotide being capable  
10 of inducing homologous recombination at the target region of the native nucleic acid; and

b) a donor nucleic acid which comprises a nucleic acid sequence sufficiently homologous to the native nucleic acid segment such that the donor nucleic acid will undergo  
15 homologous recombination with the native sequence at the target region when the third strand is bound to the native nucleic acid.

38. The composition of claim 37, wherein the  
20 oligonucleotide third strand is from about 7 to about 30 nucleotides in length.

39. The composition of claim 37, wherein the  
25 oligonucleotide third strand contains an at least partially artificial backbone.

40. The composition of claim 37, wherein the  
30 oligonucleotide third strand contains a backbone selected from the group consisting of phosphodiester, phosphorothioate, methyl phosphonate and peptide.

41. The composition of claim 40, wherein the backbone is phosphodiester.

35 42. The composition of claim 37, wherein the 3' and 5' ends of the oligonucleotide third strand are capped with one or more protective groups.

43. The composition of claim 42, wherein the protective group is selected from the group consisting of alkyl amines, thiols, cholesterol, acridine and psoralen.

5

44. The composition of claim 37, wherein the oligonucleotide third strand is circularized.

45. The composition of claim 37, wherein the oligonucleotide third strand contains at least one modified sugar.

10

46. The composition of claim 37, wherein the moiety is linked to the oligonucleotide either directly or through a linker.

15

47. The composition of claim 37, wherein the moiety is selected from the group consisting of psoralen, a substituted psoralen, hydroxymethylpsoralen, mitomycin C, 1-nitrosopyrene, a nuclease, a restriction enzyme, a radionuclide, boron, and iodine.

20

48. The composition of claim 37, wherein the non-native nucleic acid is double stranded.

25

49. The composition of claim 37, wherein the non-native nucleic acid is single stranded.

50. The composition of claim 37, wherein the non-native nucleic acid is substantially homologous to the native nucleic acid in a region of about 20 bases at each end of the non-native nucleic acid.

30

51. The composition of claim 37, wherein the non-native nucleic acid is between about 16 and about 20,000 bases in length.

35

52. The composition of claim 51, wherein the non-native nucleic acid is between about 1,000 and about 3,000 bases in length.

5 53. The composition of claim 37, wherein the donor nucleic acid is contained in a packaging system.

54. The composition of claim 53, wherein the packaging system is selected from the group consisting of a  
10 DNA virus, an RNA virus, and a liposome.

55. The composition of claim 37, wherein the native nucleic acid contains a mutation that is corrected by the recombination.

15 56. The composition of claim 55, wherein the mutation is selected from the group consisting of base changes, deletions, insertions, and combinations thereof.

20 57. A kit for effecting homologous recombination, comprising packaging material and:

a) an oligonucleotide third strand which comprises a base sequence which is capable of forming a triple helix at a binding region on one or both strands of a native nucleic  
25 acid segment in the vicinity of the human  $\beta$  globin gene cluster target region, said oligonucleotide being capable of inducing homologous recombination at the target region of the native nucleic acid; and

b) a donor nucleic acid which comprises a nucleic  
30 acid sequence sufficiently homologous to the native nucleic acid segment such that the donor nucleic acid will undergo homologous recombination with the native sequence at the target region when the third strand is bound to the native nucleic acid.

35 58. A method for site directed mutagenesis of a native nucleic acid segment in a cell, which comprises:



a) introducing into a cell an oligonucleotide third strand which comprises a base sequence which is capable of forming a triple helix at a binding region on one or both strands of a native nucleic acid segment which contains an undesired mutation in the vicinity of the human  $\beta$  globin gene cluster target region, said oligonucleotide having incorporated therein a mutagen;

b) allowing the oligonucleotide to bind to the native nucleic acid segment to form a triple stranded nucleic acid; and

c) allowing mutagenesis to occur in the target region.

59. The method of claim 58, wherein the mutagen is selected from the group consisting of psoralen, a substituted psoralen, hydroxymethylpsoralen, mitomycin C, 1-nitrosopyrene, a nuclease, a restriction enzyme, a radionuclide, boron, and iodine.

60. The method of claim 59, wherein the mutagen is psoralen.

1/3

```

      T=A
G  ** G=C
A  ** A=T
G  ** G=C
G  ** G=C
      mistake
A  ** A=T      UV-light      in repair
psoralen-A  ** A=T -----> A+T -----> T=A
      C=G
      A=T
      A=T
      T=A
```

FIGURE 1

FIGURE 2(A)

Region of  
 human chromosome:
 third strand  
binding site

-----x-----[pppppppppppp]-----

x = defective site to be repaired  
 p = purines of third strand  
 binding site

FIGURE 2(B)

M-tttttttttttt  
 -----x-----d[pppppppppppp]-----

damage
 M = mutagen capable of causing DNA

purines
 t = third strand complementary to

d = site of DNA damage

FIGURE 2(C)

-----x-----d[pppppppppppp]-----  
 D: -----[ppp

D is a donor DNA with nearly identical sequence, except for genetic defect x, of the region of the human chromosome to be replaced.

FIGURE 2(D)

New chromosome region: -----[pppppppppppp]-----

3/3

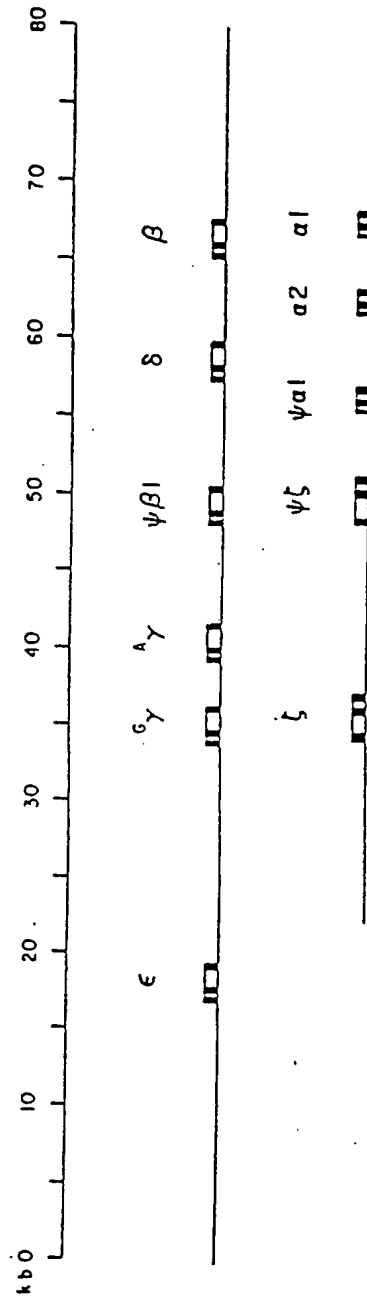


FIGURE 3

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/09430

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :A61K 48/00; C07H 21/04; C12N 5/06; 15/00,15/11,15/85, 15/86

US CL :435/172.3, 240.2, 240.21; 536/23.1, 24.1, 24.2; 514/44

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/172.3, 240.2, 240.21; 536/23.1, 24.1, 24.2; 514/44

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MOLECULAR AND CELLULAR BIOLOGY, Volume 15, Number 3, issued March 1995, G. Wang et al., "Targeted mutagenesis in mammalian cells mediated by intracellular triple helix", pages 1759-1768, see entire document.	1-60
Y	NUCLEIC ACIDS RESEARCH, Volume 22, Number 14, issued 1994, F.P. Gasparro et al., "Site-specific targeting of psoralen photoadducts with a triple helix-forming oligonucleotide: characterization of psoralen monoadduct and crosslink formation", pages 2845-2852, see entire document.	1-60
Y	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, USA, Volume 88, Number 10, issued May 1991, E.G. Shesely et al., "Correction of a human beta S-globin gene by gene targeting", pages 4294-4298, see entire document.	1-60

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*-&* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

12 AUGUST 1996

Date of mailing of the international search report

04 SEP 1996

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRIAN R. STANTON

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/09430

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, USA, Volume 92, Number 3, issued 31 January 1995, D.F. Segal et al., "Endonuclease-induced, targeted homologous extrachromosomal recombination in <i>Xenopus</i> oocytes", pages 806-810, see entire document.	1-60

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US96/09430

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

Databases: APS, Aidsline, aquasci, biobusiness, biosis, biotechds, caba, cancerlit, ca, caplus, ceaba, cen, cin, cjaes, cjelseview, confsci, cropb, cropu, dgene, dissabs, druglaunch, drugnl, drugu, embal, embasc, fsta, genbank, healsafe, ifipat, jicst-eplus, jpnews, lifesci, medline, ntis, ocean, phic, phin, prompt, scisearch, topline, toxlit, uspatfull

Search Terms: glazer?/au; triple; three; homolog?; recomb; globin#; sickle; hemoglobin?; gene; therapy; psoralen, hydroxymethylpsoralen; mitomyucin, nitrosopyrene, boron, iodine; liposom?; phospho?; acridin?;

1/3

```

                T=A
      G ** G=C
      A ** A=T
      G ** G=C
      G ** G=C
      A ** A=T      uv-light      mistake      in repair
psoralen-A ** A=T -----> A+T -----> T=A
      C=G
      A=T
      A=T
      T=A
  
```

FIGURE 1



FIGURE 2(A)

third strand  
 binding site  
 Region of  
 human chromosome: -----x-----[ppppppppppp]-----

x = defective site to be repaired  
 p = purines of third strand  
 binding site

FIGURE 2(B)

M-tttttttttttt  
 -----x-----d[ppppppppppp]-----  
 M = mutagen capable of causing DNA  
 damage  
 t = third strand complementary to  
 purines  
 d = site of DNA damage

FIGURE 2(C)

-----x-----d[ppppppppppp]-----  
 D: -----[ppp]

D is a donor DNA with nearly identical sequence, except for genetic defect x, of the region of the human chromosome to be replaced.

FIGURE 2(D)

New chromosome region: -----[ppppppppppp]-----

3/3

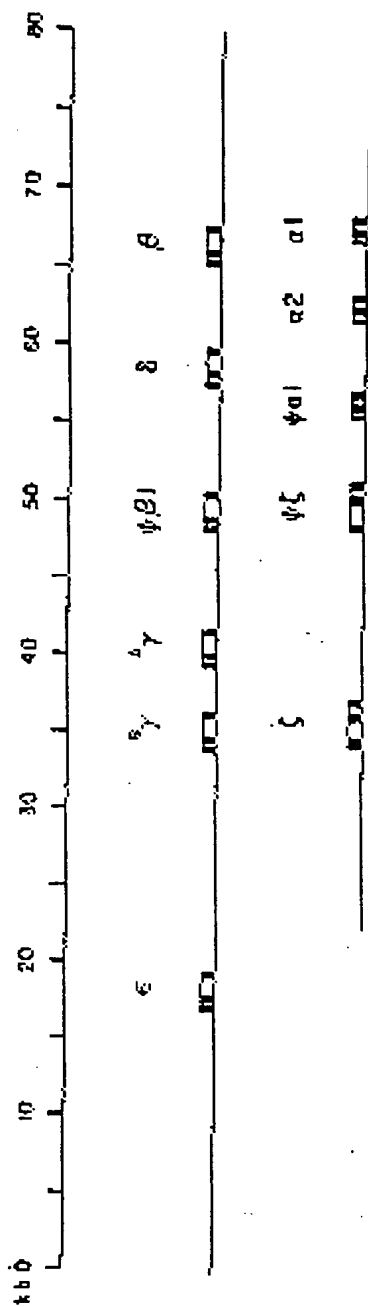


FIGURE 3